CAR-TR-840
CS-TR-3691

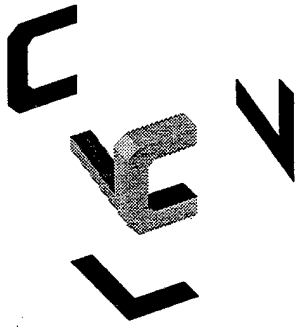
N00014-96-1-0587
September 1996

Algorithm-Independent Stability Analysis of Structure from Motion

Cornelia Fermüller and Yiannis Aloimonos

Computer Vision Laboratory
Center for Automation Research
Institute for Advanced Computer Studies
and the Computer Science Department
University of Maryland
College Park, MD 20742-3275

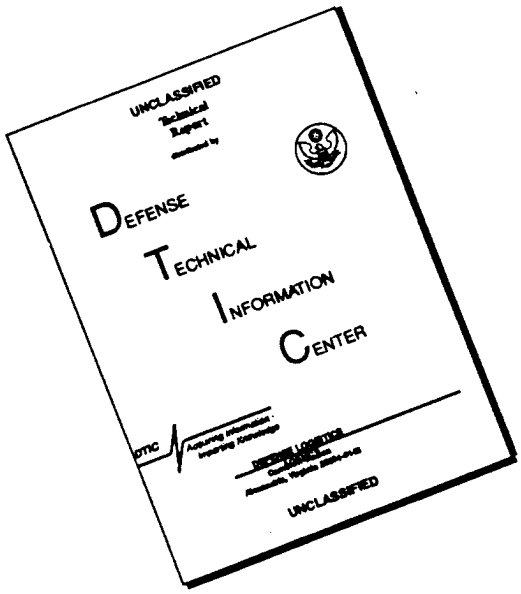
COMPUTER VISION LABORATORY



CENTER FOR AUTOMATION RESEARCH

UNIVERSITY OF MARYLAND
COLLEGE PARK, MARYLAND
20742-3275

DISCLAIMER NOTICE



THIS DOCUMENT IS BEST QUALITY AVAILABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

CAR-TR-840 N00014-96-1-0587
CS-TR-3691 September 1996

Algorithm-Independent Stability Analysis of Structure from Motion

Cornelia Fermüller and Yiannis Aloimonos

Computer Vision Laboratory
Center for Automation Research
Institute for Advanced Computer Studies
and the Computer Science Department
University of Maryland
College Park, MD 20742-3275

Abstract

The stability analysis for the structure from motion problem presented in this paper investigates the optimal relationship between the errors in the estimated translational and rotational parameters of a rigid motion that results in the estimation of a minimum number of negative depth values. No particular estimators are used and no specific assumptions about the scene are made. The input used is the value of the flow along some direction, which is more general than optic flow or correspondence. For a planar retina it is shown that the optimal configuration is achieved when the projections of the translational and rotational errors on the image plane are perpendicular. For a spherical retina, given a rotational error, the optimal translation is the correct one, while given a translational error the optimal rotational error is normal to the translational one at an equal distance from the real and estimated translations. The proofs, besides illuminating the confounding of translation and rotation in structure from motion, have an important application to ecological optics. The same analysis provides a computational explanation of why it is much easier to estimate self-motion in the case of a spherical retina and why it is much easier to estimate shape in the case of a planar retina, thus suggesting that nature's design of compound eyes (or panoramic vision) for flying systems and camera-type eyes for primates (and other systems that perform manipulation) is optimal.

The support of the Office of Naval Research under Grant N00014-96-1-0587 is gratefully acknowledged.

1 Introduction

The general problem of structure from motion is defined as follows: given a number of views of a scene, to recover the rigid transformations between the views and the structure (shape) of the scene in view. In the field of computational vision a lot of effort has been devoted to this problem because it lies at the heart of several applications in pose estimation, recognition, calibration, and navigation [8, 17]. For reasons related to the tractability of the exposition and without loss of generality, we consider here the case of differential motion for a camera moving in a static environment with the goal of recovering the camera's 3D rigid motion and the structure of the scene [42, 4, 30]. The problem has been traditionally treated in a two-step approach. The first step attempts to establish the correspondence between successive image frames, i.e., to identify in consecutive images features that are the projections of the same feature in the 3D scene. Such correspondence is expressed through displacement vectors or optic flow—an approximation of the motion field which represents the projection of the velocity field of scene points on the image. The second step attempts to interpret this correspondence or flow field and recover 3D motion and structure.

During the Eighties, questions related to the uniqueness of solutions were answered for both the discrete case of point matches [24, 41] and the differential case [25, 44]. This work gave rise to closed form solutions and opened avenues into the study of uniqueness issues. Similar problems were solved in the photogrammetric literature [35]. The algorithms developed during this phase of research were based on two frames (or views) and the use of point features. Algorithms for the case of three (or multiple) frames were introduced in [40] with the formulation of the trilinear constraints and were generalized in [10] using geometric algebra. At the same time, algorithms appeared that made use of line correspondences [39], as well as algorithms that used both point and line correspondences, thanks to the trilinear constraints [40]. Also, these results were generalized to the case of uncalibrated cameras, a situation in which only projective (or, under some assumptions, affine) structure can be recovered [9, 15, 28, 21].

The promise of the uniqueness studies gave rise to an exciting quest for practical and robust algorithms for recovering 3D structure and motion from image sequences, but this was soon to be frustrated by high sensitivity to noise in the input used (optic flow or correspondence). While many solutions have been proposed, they become problematic in the case of realistic scenes and most of them degrade ungracefully as the quality of the input deteriorates. This has motivated research on the stability of the problem; [7] contains an excellent survey of existing error analyses. We will discuss the most important results in Section 3 in more technical detail after some mathematical prerequisites are given in Section 2. In summary, it can be concluded that the majority of the existing analyses attempt to model the errors in either the 3D motion estimates or the depth estimates, and due to the large number of unknowns in the problem, they deal with restricted conditions such as planarity of the scene in view or non-biasedness of the estimators. Notably absent in published efforts is an account of the systematic nature of the errors in the depth estimates due to errors in the 3D motion estimates. Put in different terms, there exists an interplay between 3D motion and depth. In existing approaches, however, the highly correlated nature of the depth errors at different image

locations, due to 3D motion errors, is not reflected adequately. Furthermore, all analyses are based on the two-step approach, analyzing the estimation of 3D motion from noise-contaminated optic flow or correspondence. However, as has been shown in previous work, the estimation of 3D motion does not necessarily require the prior computation of exact correspondence [11, 12, 13, 20, 29]. Flow measurements, or even their signs, along some direction in the image, such as—for example—the one provided by the spatial gradient, are sufficient for recovering 3D motion [3]. Such measurements can be computed by even the simplest systems—biological or artificial—using, for example, Reichardt detectors or equivalent energy models [32, 33, 31, 43].

In this paper an approach that is independent of any algorithm or estimator is taken. Due to the geometry of image formation any spatiotemporal representation in the image is due to the 3D motion and the structure of the scene in view. If the 3D motion can be estimated correctly, the structure can be derived correctly using the equations of image formation. However, an error in the estimation of the 3D motion will result in the computation of a distorted version of the actual scene structure. Of computational interest are those regions in space where the distortions are such that the depths become negative. Not considering any scene interpretation the only fact we know about the scene is that for it to be visible it has to lie in front of the image and thus the corresponding depth estimates have to be positive. Therefore the number of image points whose corresponding scene points would yield negative values due to erroneous 3D motion estimation should be kept as small as possible. This is the computational principle behind the error analysis presented in this paper. In particular, the following questions are studied. Assuming there is an error in the estimation of the rotational motion components, what is the error in the translational components that leads to a minimization of the negative depth values computed? Similarly, if there is an error in the translational motion estimates, which rotational error will result in the smallest number of negative depth values? The analysis is carried out for a complete field of view as perceived by an imaging sphere, and for a restricted field of view on a constrained image plane.

2 Overview and Problem Statement

2.1 Prerequisites

We consider an observer moving rigidly with translation $\mathbf{t} = (U, V, W)$ and rotation $\boldsymbol{\omega} = (\alpha, \beta, \gamma)$ in a stationary environment. Thus each scene point $\mathbf{R} = (X, Y, Z)$ measured with respect to a coordinate system $OXYZ$ fixed to the camera's nodal point O has a velocity $\dot{\mathbf{R}} = -\mathbf{t} - \boldsymbol{\omega} \times \mathbf{R}$ relative to the camera. The image formation is based on perspective projection.

If the image is formed on a plane orthogonal to the Z axis at distance f from the nodal point (see Figure 1) the image points $\mathbf{r} = (x, y, f)$ are related to the scene points $\mathbf{R}$ through equation

$$\mathbf{r} = \frac{f}{\mathbf{R} \cdot \mathbf{z}_0} \mathbf{R}$$

with $\mathbf{z}_0$ a unit vector in the direction of the Z axis and "$\cdot$" denoting the inner product.

Thus, the 2D image velocity amounts to

$$\dot{\mathbf{r}} = \frac{\mathbf{v}_{\rm tr}(\mathbf{r})}{Z} + \mathbf{v}_{\rm rot}(\mathbf{r}) = -\frac{1}{Z}(\mathbf{z}_0 \times (\mathbf{t} \times \mathbf{r})) + \frac{1}{f}(\mathbf{z}_0 \times (\mathbf{r} \times (\boldsymbol{\omega} \times \mathbf{r}))) \qquad (1)$$

where $\frac{\mathbf{v}_{\rm tr}(\mathbf{r})}{Z}$ and $\mathbf{v}_{\rm rot}(\mathbf{r})$ are the translational and rotational flow components respectively and $Z = \mathbf{R} \cdot \mathbf{z}_0$.

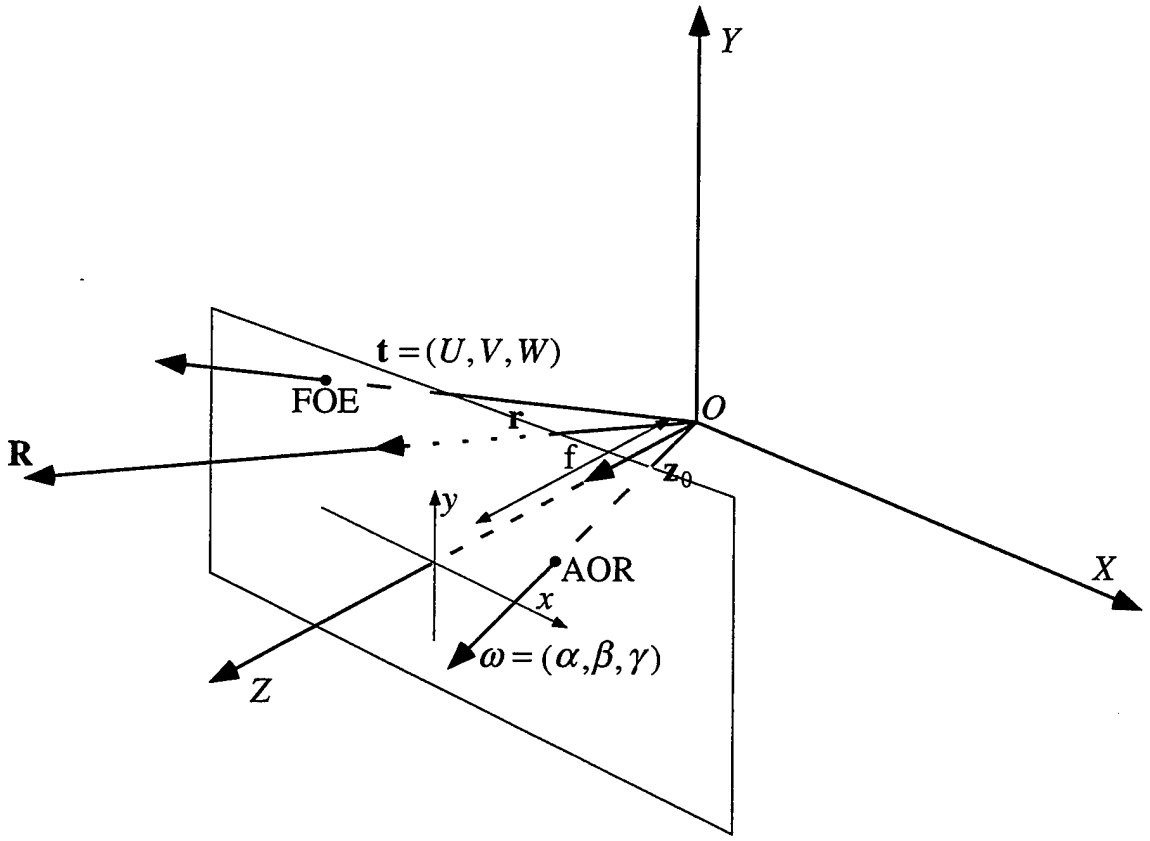


Figure 1: Image formation under perspective projection on a planar retina: The instantaneous rigid motion is described through a translation $\mathbf{t} = (U, V, W)$ and a rotation $\boldsymbol{\omega} = (\alpha, \beta, \gamma)$. The focus of expansion (FOE), given by $(\frac{U}{W}f, \frac{V}{W}f)$, denotes the direction of translation, and the AOR (axis of rotation point), given by $(\frac{\alpha}{\gamma}f, \frac{\beta}{\gamma}f)$, denotes the intersection of the rotation axis and the image.

Similarly, if the image is formed on a sphere of radius f (i.e., $\mathbf{r} \cdot \mathbf{r} = f^2$) (see Figure 2), the image $\mathbf{r} = (x, y, z)$ of any point $\mathbf{R}$ is

$$\mathbf{r} = \frac{\mathbf{R}f}{|\mathbf{R}|}$$

with $|\mathbf{R}|$ being the norm of $\mathbf{R}$ and denoting the range; thus the 2D image motion is

$$\dot{\mathbf{r}} = \frac{\mathbf{v}_{\rm tr}(\mathbf{r})}{|\mathbf{R}|} + \mathbf{v}_{\rm rot}(\mathbf{r}) = -\frac{1}{|\mathbf{R}|f}(\mathbf{r} \times (\mathbf{r} \times \mathbf{t})) - \boldsymbol{\omega} \times \mathbf{r} \qquad (2)$$

The component of the flow u_n along any direction $\mathbf{n}$ is therefore

$$u_n = \dot{\mathbf{r}} \cdot \mathbf{n} = \frac{\mathbf{v}_{\text{tr}}}{Z} \cdot \mathbf{n} + \mathbf{v}_{\text{rot}} \cdot \mathbf{n} \quad \text{or} \quad u_n = \dot{\mathbf{r}} \cdot \mathbf{n} = \frac{\mathbf{v}_{\text{tr}}}{|\mathbf{R}|} \cdot \mathbf{n} + \mathbf{v}_{\text{rot}} \cdot \mathbf{n} \tag{3}$$

As can be seen from equations (1) and (2), the effects of translation and scene structure cannot be disentangled and thus we can only obtain the direction of translation $\mathbf{t}/|\mathbf{t}|$ and the depth (range) of the scene up to a scaling factor, that is $\frac{Z}{|\mathbf{t}|}\left(\frac{|\mathbf{R}|}{|\mathbf{t}|}\right)$. For the sake of simplicity, we will assume $\mathbf{t}$ to be of length 1 and we will no longer mention the scaling in the computation of structure.

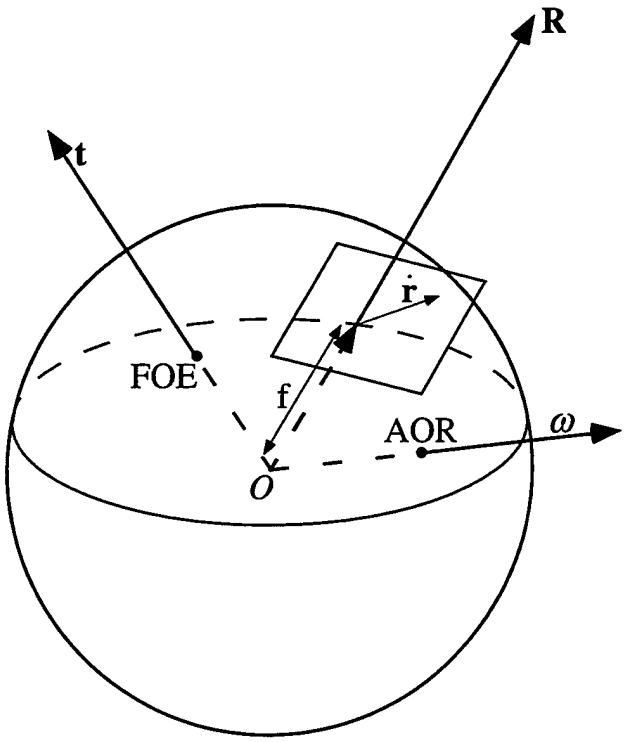


Figure 2: Image formation under perspective projection on a spherical retina.

2.2 Previous work

It is in general a very hard task to develop analytical results about the stability or error sensitivity of structure from motion. This is due to the nonlinearities and the large number of parameters that are involved. As a result a fair number of observations and intuitive arguments have been developed by a multitude of authors over the years. Most important, a small number of studies have given rise to three crisp results regarding noise sensitivity in structure from motion [7]. These are:

(a) A translation can be easily confounded with a rotation in the case of a small field of view under the assumption of lateral motion and insufficient variation of depth

[1, 6]. Intuitively, translation along the x axis can be confused with rotation around the y axis and translation along the y axis with rotation around the x axis. Evidence for this result can be obtained intuitively from the flow equation (1). As can be seen, if the scene in view is a plane, then the flow becomes a polynomial in the retinal coordinates x, y with the terms $\mathbf{t}_1 + \boldsymbol{\omega}_2$, $\mathbf{t}_2 - \boldsymbol{\omega}_1$ representing the zero-order terms. An elegant proof of this fact using techniques from estimation theory has been presented by Daniilidis [6] for the case of unbiased estimators.

(b) Usually an error metric is developed whose minimization provides a solution for 3D motion and subsequently for structure. If this metric is not appropriately normalized, in the case of a small field of view the translation estimate is biased toward the viewing direction. This can be seen directly from the epipolar constraint. In its instantaneous form the epipolar constraint becomes $(\mathbf{t} \times \mathbf{r}) \cdot (\dot{\mathbf{r}} - \boldsymbol{\omega} \times \mathbf{r}) = 0$ assuming $f = 1$. In the discrete case, if $\mathbf{r}_1$ and $\mathbf{r}_2$ are corresponding image points before and after the motion the constraint is $\mathbf{r}_2 \cdot (\mathbf{t} \times R\mathbf{r}_1) = 0$, where R represents the rotation matrix. A solution coming from the minimization of $\sum_i \|\mathbf{r}_{2i} \cdot (\mathbf{t} \times R\mathbf{r}_{1i})\|^2$ is bound to be biased, because the cross product introduces the sine of the angle between vectors $\mathbf{t}$ and $R\mathbf{r}_{1i}$ as a factor. This, in turn, makes the minimization prefer vectors $\mathbf{t}$ that are closer to the center of gravity of the points $R\mathbf{r}_{1i}$ so that the sine and hence the residual is smaller [36, 38]. Techniques from statistics such as maximum likelihood estimation [36] or Rayleigh optimization [37] can be used to deal with the bias, but they have their own problems.

(c) The third result is due to Maybank [26, 27] who showed that in the case of a small field of view and an irregular surface the cost function resulting from the epipolar constraint, $\sum_i \|(\mathbf{t} \times \mathbf{r}_i) \cdot (\dot{\mathbf{r}}_i - \boldsymbol{\omega} \times \mathbf{r}_i)\|^2$, takes its minima along a line in the space of translation directions which passes through the true translation and (not surprisingly due to the small field of view assumption) the viewing direction. This means that the tilt of the direction of $\mathbf{t}$ can be estimated more reliably than its slant.

Additional important work has been concerned with the study of configurations of scene points that give rise to multiple solutions from point correspondences [18], the so-called ambiguity-critical surfaces. It has been shown by Horn [19] that the epipolar constraint is not affected by first-order deformations of the motion parameters if the points lie on a quadric surface with certain properties. The relationship between these instability-critical surfaces and the ambiguity-critical surfaces has been established in [6, 16].

Next, we study the relationship between errors in the estimation of the 3D motion and errors in the estimation of the depth of the scene. This relationship is the basis for our subsequent error analysis.

2.3 Distorted space

Based on an exact computation of the motion parameters the depth (range) can be derived from equation (3). Let us assume, however, there is an error in the estimation of the five motion parameters, that is the two parameters of the direction of translation

and the three parameters of rotation. As a consequence there will also be errors in the estimation of depth (range) and thus a distorted version of the space will be computed. A convenient way to describe the distortion of space is to sketch it through surfaces in space which are distorted by the same multiplicative factor, the iso-distortion surfaces [5, 14].

In the following, in order to distinguish between the various estimates, we use letters with hat signs to represent the estimated quantities $(\hat{\mathbf{t}}, \hat{\omega}, |\hat{\mathbf{R}}|, \hat{Z}, \hat{\mathbf{v}}_{\text{tr}}, \hat{\mathbf{v}}_{\text{rot}})$ and unmarked letters to represent the actual quantities $(\mathbf{t}, \omega, |\mathbf{R}|, Z, \mathbf{v}_{\text{tr}}, \mathbf{v}_{\text{rot}})$. The subscript "$\epsilon$" is used to denote errors, where we define $\omega - \hat{\omega} = \omega_\epsilon$ and $\mathbf{v}_{\text{rot}} - \hat{\mathbf{v}}_{\text{rot}} = \mathbf{v}_{\text{rot}_\epsilon}$.

The estimated depth or range can be obtained from equation (3) as

$$\hat{Z} \text{ (or } |\hat{\mathbf{R}}|) = \frac{\hat{\mathbf{v}_{\text{tr}}} \cdot \mathbf{n}}{\dot{\mathbf{r}} \cdot \mathbf{n} - \hat{\mathbf{v}}_{\text{rot}} \cdot \mathbf{n}}$$

and we have on the image plane

$$\hat{Z} = Z \left(\frac{-f \left(\mathbf{z}_0 \times \left(\hat{\mathbf{t}} \times \mathbf{r} \right) \right) \cdot \mathbf{n}}{-f \left(\mathbf{z}_0 \times (\mathbf{t} \times \mathbf{r}) \right) \cdot \mathbf{n} + Z \left(\mathbf{z}_0 \times (\mathbf{r} \times (\omega_\epsilon \times \mathbf{r})) \right) \cdot \mathbf{n}} \right) \tag{4}$$

and on the image sphere

$$|\hat{\mathbf{R}}| = |\mathbf{R}| \cdot \left(\frac{\left(\mathbf{r} \times \left(\mathbf{r} \times \hat{\mathbf{t}} \right) \right) \cdot \mathbf{n}}{(\mathbf{r} \times (\mathbf{r} \times \mathbf{t})) \cdot \mathbf{n} + f |\mathbf{R}| \left(\omega_\epsilon \times \mathbf{r} \right) \cdot \mathbf{n}} \right) \tag{5}$$

From equation (4) it can be seen that $\hat{Z}$ can be expressed as a multiple of Z, where the multiplicative factor, which we denote by D, the distortion factor, is given by the term inside the brackets. Thus the distortion factor is

$$D = \frac{-f \left(\mathbf{z}_0 \times \left(\hat{\mathbf{t}} \times \mathbf{r} \right) \right) \cdot \mathbf{n}}{-f \left(\mathbf{z}_0 \times (\mathbf{t} \times \mathbf{r}) \right) \cdot \mathbf{n} + Z \left(\mathbf{z}_0 \times (\mathbf{r} \times (\omega_\epsilon \times \mathbf{r})) \right) \cdot \mathbf{n}} \tag{6}$$

Similarly we can interpret the estimated range in equation (5) as a multiple of the actual range with distortion D, where

$$D = \frac{\left(\mathbf{r} \times \left(\mathbf{r} \times \hat{\mathbf{t}} \right) \right) \cdot \mathbf{n}}{(\mathbf{r} \times (\mathbf{r} \times \mathbf{t})) \cdot \mathbf{n} + f |\mathbf{R}| \left(\omega_\epsilon \times \mathbf{r} \right) \cdot \mathbf{n}} \tag{7}$$

Equations (6) and (7) describe, for any fixed direction $\mathbf{n}$ and any distortion factor D, a surface in space. Any such surface is to be understood as the locus of points in space which are distorted in depth (range) by the same factor D, if the corresponding image measurements are in direction $\mathbf{n}$.

It should be emphasized that the distortion of depth also depends on the direction $\mathbf{n}$ of the flow measurement (hereafter called the flow direction) used as basis for the computations and therefore is different for different directions of flow. This means simply that if one estimates depth from optical flow in the presence of errors, the results can be

very different, depending on whether the horizontal, vertical, or any other component is used. Depending on the direction, any value between $+\infty$ and $-\infty$ can be obtained!

In the analysis in this paper, we are not interested in actual 3D space, but we consider the surfaces in visual space, that is, the space perceived under perspective projection where the dimensions parallel to the image are measured according to the size with which they appear on the image.

Figure 3a gives an example of an iso-distortion surface, and Figure 3b illustrates a family of iso-distortion surfaces corresponding to the same gradient direction but different distortion factors D. The same family is intersected with the xZ plane in Figure 3c. In the plane the intersections give rise to a family of contours.

As can be seen the iso-distortion surfaces of a family intersect in a curve, and they change continuously as we vary D. Thus all the space between the 0 distortion surface and the $-\infty$ distortion surface (which is also the $+\infty$ distortion surface) is distorted by a negative distortion factor.

2.4 Description of results

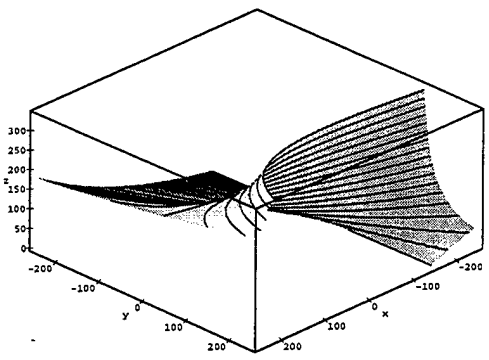
In the forthcoming sections we employ a geometric statistical model to represent the negative depth values. We assume that the scene in view lies within a certain depth (range) interval between a minimum value and a maximum value. The flow representation vectors in the image are in different directions, and we assume some distribution for their directions. Our focus is on the points in space which for a 3D motion estimate yield negative depth (range) estimates.

For every direction $\mathbf{n}$ the points in space with negative depth estimates cover the space between the 0 and $-\infty$ distortion surface within the range covered by the scene. Thus for every direction we obtain a 3D subspace, covering a certain volume. The sum of all volumes for all directions, normalized by the flow distributions considered, represents a measure of the likelihood that negative depth values occur. We call it the "negative depth volume" or "negative range volume." The idea behind our error analysis lies in the minimization of this negative depth (range) volume—that is, we are interested in the relationship between the translational and rotational motion errors that minimizes this volume.

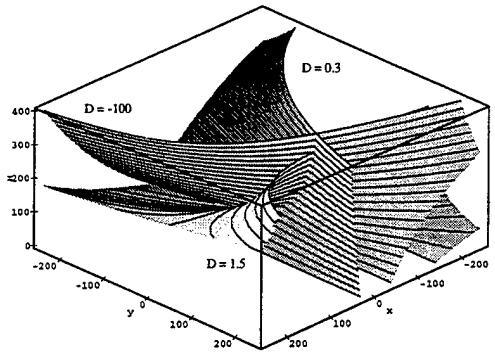
In our analysis we do not want to make any particular scene-related assumptions favoring particular orientations or depth values. We wish to treat all depth values and flow directions as having equal importance. To be more precise, we assume that the flow directions are uniformly distributed in every direction and at every depth (range) between a minimum value $Z_{\min}(|\mathbf{R}_{\min}|)$ and a maximum value $Z_{\max}(|\mathbf{R}_{\max}|)$. We do not wish to assume any particular distribution for the noise in the flow measurements. Therefore, we do not consider any noise in the measurements. Thus, one can view our analysis as a geometric investigation of the inherent confounding of translation and rotation, which is the reason behind the instability in structure from motion.

In summary, as an answer to the question about the coupling of motion errors, the following results are obtained:

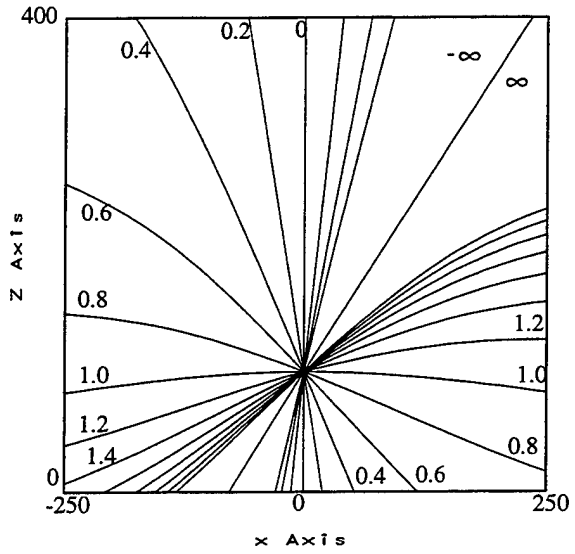
(a) If we take the whole sphere as the imaging surface and we assume an error in the



(a)



(b)



(c)

Figure 3: (a) Iso-distortion surface in xyZ space. The parameters are: $(x_0, y_0) = (\frac{U}{W}f, \frac{V}{W}f) = (-50, -25)$, $(\hat{x}_0, \hat{y}_0) = (\frac{\hat{U}}{\hat{W}}f, \frac{\hat{V}}{\hat{W}}f) = (0, -20)$, $\omega_\epsilon = (\alpha_\epsilon, \beta_\epsilon, \gamma_\epsilon) = (-0.005, 0.001, 0.003)$, $D = 1.5$, $\mathbf{n} = (1, 0)$, $f = 500$ (corresponding to a field of view of 50°). (b) Family of iso-distortion surfaces for the same motion parameters ($\mathbf{n} = (1, 0)$). (c) Corresponding iso-distortion contours in the xZ plane.

estimation of rotation, then the direction of translation that minimizes the negative depth volume is the correct direction of translation.

The practical implication of this result is that 3D motion estimation is most easily accomplished for a complete field of view, as provided by an imaging sphere. A working system (biological or artificial) is usually equipped with an inertial sensor which provides rotational information, though probably with some error. On the basis of this information, the best one can do to estimate the remaining translation is to assume that the flow field obtained by subtracting the estimated rotation is purely translational and apply a simple algorithm designed for only translation [2, 20, 29, 34].

Such algorithms, if based only on the constraint that the depth is positive, are formulated basically as constrained minimization problems. The underlying idea is illustrated in Figure 4. Assuming the observer is approaching the scene, the exact 2D motion vector at every point is away from the FOE (the point where the translation axis pierces the image). Thus the projection $\mathbf{u}_n$ of the flow vector on any direction $\mathbf{n}$ is confined to lie in a half-plane, as defined by line ϵ in Figure 4, and the FOE is to be found in the complementary half-plane. Thus the estimation of the translational direction can be implemented by simply voting for a half-plane at every point. The best solution corresponds to the location with the highest number of votes.

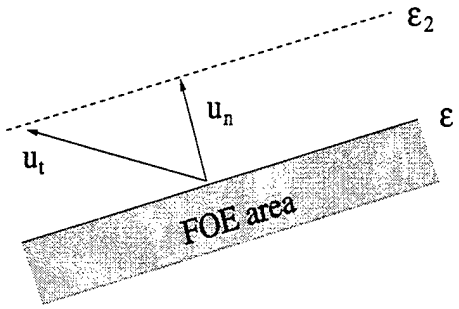


Figure 4: The translational flow vector $\boldsymbol{u_t}$ has its tip anywhere along the line ε_2. The focus of expansion lies on the (shaded) half plane defined by the line ε that does not contain possible vectors $\boldsymbol{u_t}$.

Estimation of purely translational motion is much simpler than estimation of complete 3D rigid motion, which requires techniques that decouple the translation from the rotation in some way, and if designed on the basis of the constraint of positive depth, require voting in higher dimensions [11, 12, 13].

As demonstrated in the forthcoming analysis, however, a simple algorithm designed for translation only will find the correct solution. Thus insects with spherical eyes, such as bees and flies, have a big advantage in the task of 3D motion estimation.

(b) On the other hand, if we assume a certain error in the estimation of translation on a spherical image, we find that the vector of the rotational error $\boldsymbol{\omega}_\epsilon$ lies on the same

geodesic as the real translation $\mathbf{t}$ and the estimated translation $\hat{\mathbf{t}}$ at equal distance from both, that is, $(\mathbf{t} + \hat{\mathbf{t}}) \times \boldsymbol{\omega}_\epsilon = 0$ ($\mathbf{t}$ and $\hat{\mathbf{t}}$ are unit vectors).

(c) Considering as imaging surface a plane of limited extent, we find that the translational and rotational errors are perpendicular to each other. Using the notation $\frac{Uf}{W} - \frac{\hat{U}f}{\hat{W}} = x_{0_\epsilon}$ and $\frac{Vf}{W} - \frac{\hat{V}f}{\hat{W}} = y_{0_\epsilon}$, this means that $\frac{x_{0_\epsilon}}{y_{0_\epsilon}} = -\frac{\beta_\epsilon}{\alpha_\epsilon}$. If we fix the rotational error $(\alpha_\epsilon, \beta_\epsilon, \gamma_\epsilon)$, this provides us with a constraint on the direction of the translational error.

(d) If we fix the translational error $(x_{0_\epsilon}, y_{0_\epsilon})$ we obtain the same constraint, and in addition we find that $\gamma_\epsilon = 0$.

The results developed in this paper have a clear relationship with those of existing error analyses as described in Section 2.2, with the exception of the bias of translation towards the viewing direction, since this result has been obtained on the basis of particular algorithms and image measurement configurations.

Regarding the confusion between translation and rotation, it has been experimentally observed and proven for simple scene structures, restricted fields of view, and certain estimation techniques using particular statistical estimators, that the translation along the x axis is coupled with rotation around the y axis and that the translation along the y axis is coupled with rotation around the x axis. This can be explained using the constraint we have developed. If β_ϵ changes, the constraint $\frac{x_{0_\epsilon}}{y_{0_\epsilon}} = -\frac{\beta_\epsilon}{\alpha_\epsilon}$ remains intact if x_{0_ϵ} is changed appropriately. Similarly, an error in α_ϵ can be hidden in y_{0_ϵ}. This, however, is not true in general. An error in β_ϵ could be coupled with an error in y_{0_ϵ} or in both x_{0_ϵ} and y_{0_ϵ}. The only condition that must be satisfied is the perpendicularity between the translational and rotational errors; the confusions between x-translation and y-rotation, and y-translation and x-rotation, are not decoupled.

Regarding the distribution of the global minima of the objective function derived from the epipolar constraint, there is an interesting connection. Given a particular rotational error, our result shows perpendicularity of the translational and rotational errors irrespective of the scene in view. Thus, all possible estimated translations are to be found on a line passing through the real translation normal to the rotational error. In the analysis of Maybank, various assumptions are made that can be interpreted as introducing a rotational error. Therefore, the line found in [26, 27] should be perpendicular to the rotational error due to these assumptions.

The importance of the results obtained for the plane also lies in their consequences for shape estimation. They can be translated into the statement that planar retinas with high resolution at the center are advantageous in the computation of shape. As will be shown in Section 5, if $\frac{x_{0_\epsilon}}{y_{0_\epsilon}} = -\frac{\beta_\epsilon}{\alpha_\epsilon}$, near the fixation center for any depth Z, the distortion factor is approximately the same for every flow direction! This means that all scene points of the same depth are distorted by the same factor and thus a depth map is derived whose level contours are the correct ones!

3 Analysis on the Sphere

3.1 Fixed rotational error

We need a parameterization for expressing all possible orientations $\mathbf{n}$ tangent to the sphere at every point. One way to achieve this that is convenient for our problem is through the selection of an arbitrary unit vector $\mathbf{s}$. Given $\mathbf{s}$, at each point $\mathbf{r}$ of the sphere, the vector $\frac{\mathbf{s}\times\mathbf{r}}{|\mathbf{s}\times\mathbf{r}|}$ defines a direction at the tangent plane. As $\mathbf{s}$ varies along half a great circle, $\frac{\mathbf{s}\times\mathbf{r}}{|\mathbf{s}\times\mathbf{r}|}$ takes on every possible orientation in the tangent plane at every point $\mathbf{r}$ with the exception of the set of points $\mathbf{r}$ lying on the great circle of $\mathbf{s}$, which is of measure zero. To facilitate the analysis, we choose $\mathbf{s}$ perpendicular to ω_ϵ.

As shown in Figure 5, let ω_ϵ be parallel to the x axis and let $\mathbf{s}$ be the set of all the unit vectors in the yz plane with $\mathbf{s} = (0, \sin\chi, \cos\chi)$ and χ in the interval $[0 \ldots \pi]$. The flow directions $\mathbf{n}$ at every point are defined as $\mathbf{n} = \frac{\mathbf{s}\times\mathbf{r}}{|\mathbf{s}\times\mathbf{r}|}$. This parameterization, however, does not treat all orientations equally (as $\mathbf{s}$ varies along a great circle with constant speed, $\mathbf{s}\times\mathbf{r}$ accelerates and decelerates). Thus, in order to obtain a uniform distribution we must perform some normalization. Luckily, however, this normalization does not complicate matters in the following proof because, due to symmetry, its behavior with regard to monotonicity is the same as the of one of the volumes of negative depth in space for the functions considered.

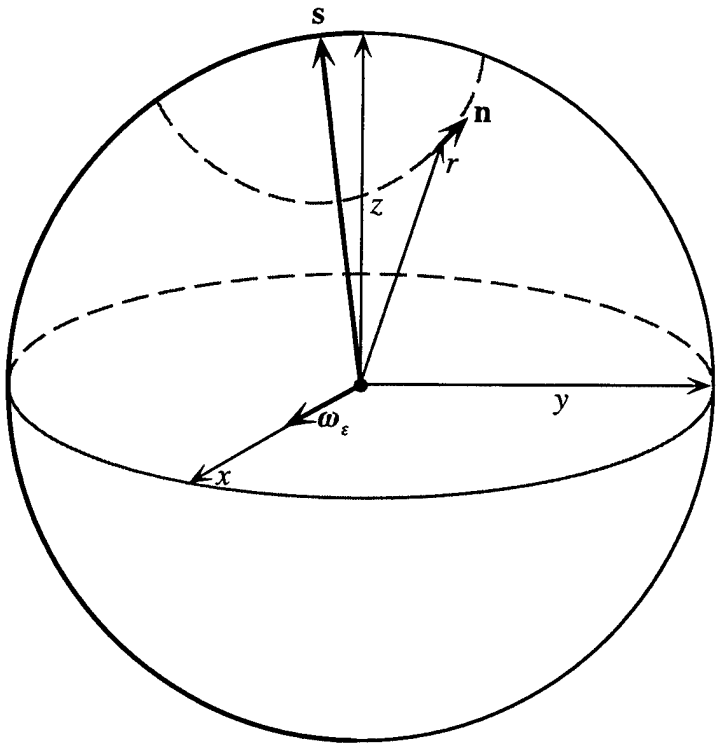


Figure 5: Parameterization used in the analysis: $\omega_\epsilon = \lambda(1,0,0)$, $\mathbf{s} = (0, \sin\chi, \cos\chi)$ with $\chi \in [0 \ldots \pi]$, $\mathbf{n} = \frac{\mathbf{s}\times\mathbf{r}}{|\mathbf{s}\times\mathbf{r}|}$.

We assume a uniform distribution for the directions $\mathbf{n}$. Thus, in order to obtain the negative range volume V_n, we have to integrate the individual volumes in each direction over all directions. If $\psi \in [0, \pi]$ provides a uniform parameterization for $\mathbf{n}$, as given in Appendix A, $V(\psi)$ is the volume for a single direction $\mathbf{n}(\psi)$, and χ is the parameterization for $\mathbf{n}$ as defined above, the following transformation applies:

$$V_n = \int_0^{\pi} V(\psi)\, d\psi = \int_{g^{-1}(0)}^{g^{-1}(\pi)} V\left(g(\chi)\right) \left| \frac{\partial g(\chi)}{\partial \chi} \right| d\chi$$

where $\psi = g(\chi)$. For this parameterization the normalization term is

$$\left| \frac{\partial g(\chi)}{\partial \chi} \right| = \left| \frac{\sin \varphi_y}{\cos(\varphi_y)^2 \cos(\chi - \varphi_x)^2 - 1} \right| \tag{8}$$

where φ_y is the angle between vector $\mathbf{r}$ and the yz plane, and φ_x is the angle between the projection of $\mathbf{r}$ on the yz plane and some fiducial direction in the yz plane. A derivation is given in Appendix A.

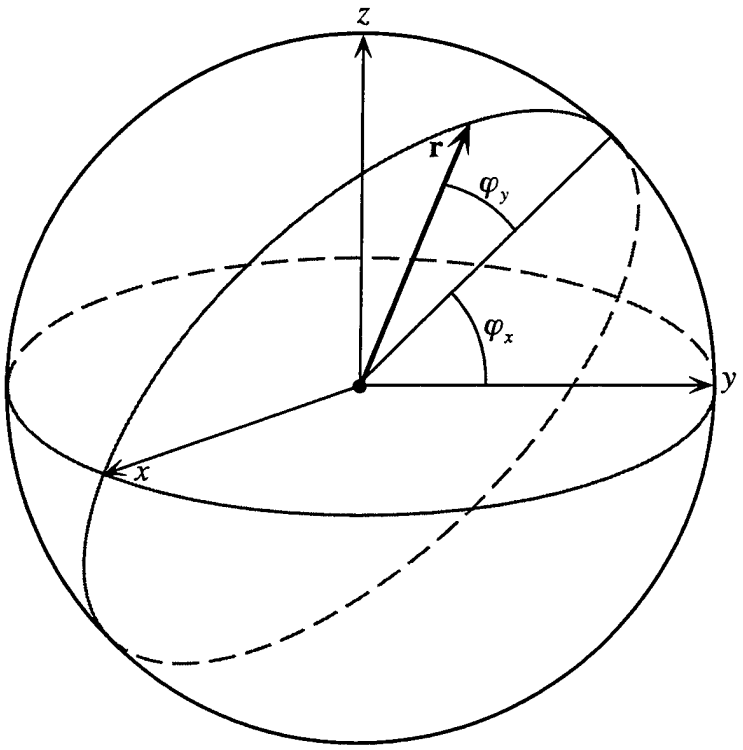


Figure 6: Parameterization of $\mathbf{r}$: φ_y is the angle between $\mathbf{r}$ and the yz plane; φ_x is the angle between the projection $\mathbf{r}$ on the yz plane and some direction in the yz plane.

Our focus is on the points in space with estimated negative range values $|\hat{\mathbf{R}}|$. Since $\mathbf{n} = \frac{\mathbf{s} \times \mathbf{r}}{|\mathbf{s} \times \mathbf{r}|}$ and $\mathbf{s} \cdot \boldsymbol{\omega}_\epsilon = 0$, we obtain from equation (5), by setting $f = 1$,

$$\left|\hat{\mathbf{R}}\right| = |\mathbf{R}| \frac{\left(\hat{\mathbf{t}} \times \mathbf{s}\right) \cdot \mathbf{r}}{(\mathbf{t} \times \mathbf{s}) \cdot \mathbf{r} - |\mathbf{R}| \left(\boldsymbol{\omega}_\epsilon \cdot \mathbf{r}\right) (\mathbf{s} \cdot \mathbf{r})} < 0 \tag{9}$$

From this inequality the following constraint on $|\mathbf{R}|$ can be derived:

$$\mathrm{sgn}(\hat{\mathbf{t}} \times \mathbf{s}) \cdot \mathbf{r} = -\mathrm{sgn}\left((\mathbf{t} \times \mathbf{s}) \cdot \mathbf{r} - |\mathbf{R}|(\boldsymbol{\omega}_\epsilon \cdot \mathbf{r})(\mathbf{s} \cdot \mathbf{r})\right) \tag{10}$$

At any point $\mathbf{r}$ in the image this constraint is either satisfied for all values $|\mathbf{R}|$, or it is satisfied for an interval of values $|\mathbf{R}|$ bounded from either above or below, or it is not satisfied for any value at all. Thus, inequality (9) provides a classification for the points on the sphere, and we obtain four different kinds of areas (types I–IV). The locations of these areas are defined by the signs of the functions $(\hat{\mathbf{t}} \times \mathbf{s}) \cdot \mathbf{r}$, $(\mathbf{t} \times \mathbf{s}) \cdot \mathbf{r}$ and $(\boldsymbol{\omega}_\epsilon \cdot \mathbf{r})(\mathbf{s} \cdot \mathbf{r})$, as summarized in Table 1.

Table 1:

area	location	constraint on $\lvert\mathbf{R}\rvert$
I	$\mathrm{sgn}(\mathbf{t} \times \mathbf{s}) \cdot \mathbf{r} = \mathrm{sgn}(\hat{\mathbf{t}} \times \mathbf{s}) \cdot \mathbf{r} = \mathrm{sgn}(\mathbf{r} \cdot \boldsymbol{\omega}_\epsilon)(\mathbf{r} \cdot \mathbf{s})$	$\lvert\mathbf{R}\rvert > \frac{(\mathbf{t} \times \mathbf{s}) \cdot \mathbf{r}}{(\mathbf{r} \cdot \boldsymbol{\omega}_\epsilon)(\mathbf{r} \cdot \mathbf{s})}$
II	$-\mathrm{sgn}(\mathbf{t} \times \mathbf{s}) \cdot \mathbf{r} = \mathrm{sgn}(\hat{\mathbf{t}} \times \mathbf{s}) \cdot \mathbf{r} = \mathrm{sgn}(\mathbf{r} \cdot \boldsymbol{\omega}_\epsilon)(\mathbf{r} \cdot \mathbf{s})$	all $\lvert\mathbf{R}\rvert$
III	$\mathrm{sgn}(\mathbf{t} \times \mathbf{s}) \cdot \mathbf{r} = -\mathrm{sgn}(\hat{\mathbf{t}} \times \mathbf{s}) \cdot \mathbf{r} = \mathrm{sgn}(\mathbf{r} \cdot \boldsymbol{\omega}_\epsilon)(\mathbf{r} \cdot \mathbf{s})$	$\lvert\mathbf{R}\rvert < \frac{(\mathbf{t} \times \mathbf{s}) \cdot \mathbf{r}}{(\mathbf{r} \cdot \boldsymbol{\omega}_\epsilon)(\mathbf{r} \cdot \mathbf{s})}$
IV	$\mathrm{sgn}(\mathbf{t} \times \mathbf{s}) \cdot \mathbf{r} = \mathrm{sgn}(\hat{\mathbf{t}} \times \mathbf{s}) \cdot \mathbf{r} = -\mathrm{sgn}(\mathbf{r} \cdot \boldsymbol{\omega}_\epsilon)(\mathbf{r} \cdot \mathbf{s})$	none

Thus for any direction $\mathbf{n}$ defined by a certain $\mathbf{s}$, we obtain a volume of negative range values consisting of the volumes above areas I, II, and III. An illustration for both hemispheres is given in Figure 7. As can be seen, areas II and III cover the same amount of area, which has the size of the area between the two great circles $(\mathbf{t} \times \mathbf{s}) \cdot \mathbf{r} = 0$ and $(\hat{\mathbf{t}} \times \mathbf{s}) \cdot \mathbf{r} = 0$, and area I covers a hemisphere minus the area between $(\mathbf{t} \times \mathbf{s}) \cdot \mathbf{r} = 0$ and $(\hat{\mathbf{t}} \times \mathbf{s}) \cdot \mathbf{r} = 0$.

If the scene in view is unbounded, that is, $|\mathbf{R}| \in [0 \ldots \infty]$, there is a range of values $|\mathbf{R}|$ above any point $\mathbf{r}$ in areas I and III which results in negative range estimates. If we consider a lower bound $|\mathbf{R}_{\min}| \neq 0$ and an upper bound $|\mathbf{R}_{\max}| \neq \infty$, we obtain two additional curves $C_{\min}$ and $C_{\max}$ with $C_{\min} = (\mathbf{t} \times \mathbf{s}) \cdot \mathbf{r} - |\mathbf{R}_{\min}|\,(\boldsymbol{\omega}_\epsilon \cdot \mathbf{r})(\mathbf{s} \cdot \mathbf{r}) = 0$ and $C_{\max} = (\mathbf{t} \times \mathbf{s}) \cdot \mathbf{r} - |\mathbf{R}_{\max}|\,(\boldsymbol{\omega}_\epsilon \cdot \mathbf{r})(\mathbf{s} \cdot \mathbf{r}) = 0$ as bounds for areas with negative range values (as shown in Figure 7). As can be seen, the curves $C_{\min} = 0$, $C_{\max} = 0$, $(\mathbf{t} \times \mathbf{s}) \cdot \mathbf{r} = 0$ and $(\boldsymbol{\omega}_\epsilon \cdot \mathbf{r})(\mathbf{s} \cdot \mathbf{r}) = 0$ intersect at the same point.

In area I, we do not obtain any volume of negative range estimates for points $\mathbf{r}$ between the curves $(\boldsymbol{\omega}_\epsilon \cdot \mathbf{r})(\mathbf{s} \cdot \mathbf{r}) = 0$ and $C_{\max} = 0$; the volume for points $\mathbf{r}$ between $C_{\min} = 0$ and $C_{\max} = 0$ is bounded from below by $|\mathbf{R}| = \frac{(\mathbf{t} \times \mathbf{s}) \cdot \mathbf{r}}{(\boldsymbol{\omega}_\epsilon \cdot \mathbf{r})(\mathbf{s} \cdot \mathbf{r})}$ (and from above by $|\mathbf{R}_{\max}|$) and the volume for points $\mathbf{r}$ between $C_{\min} = 0$ and $(\mathbf{t} \times \mathbf{s}) \cdot \mathbf{r} = 0$ extends from $|\mathbf{R}_{\min}|$ to $|\mathbf{R}_{\max}|$. In area III we do not obtain any volume for points $\mathbf{r}$ between $(\mathbf{t} \times \mathbf{s}) \cdot \mathbf{r} = 0$ and $C_{\min} = 0$. The volume for points $\mathbf{r}$ between $C_{\min} = 0$ and $C_{\max} = 0$ is bounded from above by $|\mathbf{R}| = \frac{(\mathbf{t} \times \mathbf{s}) \cdot \mathbf{r}}{(\boldsymbol{\omega}_\epsilon \cdot \mathbf{r})(\mathbf{s} \cdot \mathbf{r})}$ (and from below by $|\mathbf{R}_{\min}|$) and the volume for points $\mathbf{r}$ between $C_{\max} = 0$ and $(\boldsymbol{\omega}_\epsilon \cdot \mathbf{r})(\mathbf{s} \cdot \mathbf{r}) = 0$ extends from $|\mathbf{R}_{\min}|$ to $|\mathbf{R}_{\max}|$.

We are given $\boldsymbol{\omega}_\epsilon$ and $\mathbf{t}$, and we are interested in $\hat{\mathbf{t}}$, which minimizes the negative range volume. For any $\mathbf{s}$ the corresponding negative range volume becomes smallest if $\hat{\mathbf{t}}$ is on the great circle of $\mathbf{t}$ and $\mathbf{s}$, that is, $(\mathbf{t} \times \mathbf{s}) \cdot \hat{\mathbf{t}} = 0$, as will be shown next.

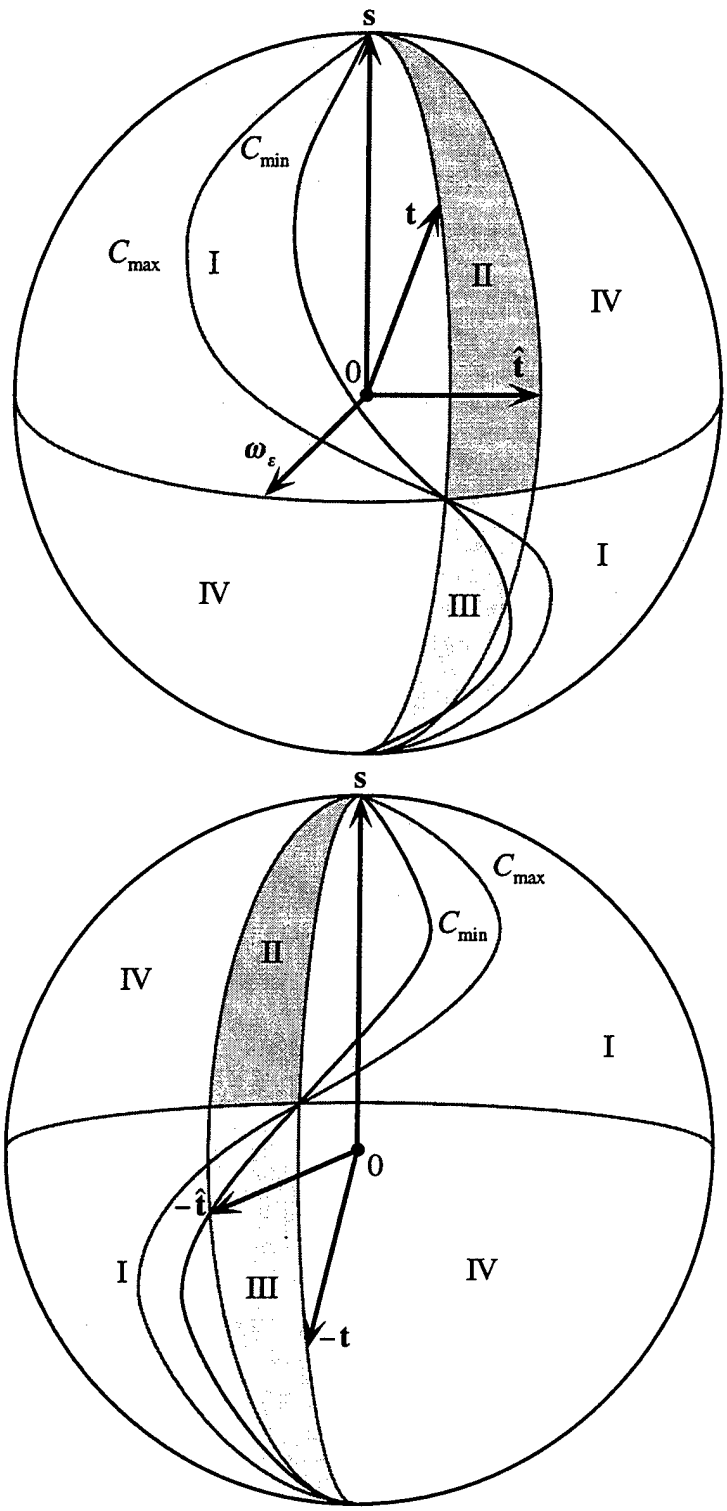


Figure 7: Classification of image points according to constraints on $|\mathbf{R}|$. At $C_{\min}$ and $C_{\max}$, $|\mathbf{R}|$ is constrained to be greater (area I) or smaller (area III) than $|\mathbf{R}_{\min}|$ or $|\mathbf{R}_{\max}|$. The two hemispheres correspond to the front of the sphere and the back of the sphere, both as seen from the front of the sphere.

Let us consider a $\hat{\mathbf{t}}$ such that $(\mathbf{t} \times \mathbf{s}) \cdot \hat{\mathbf{t}} \neq 0$ (i.e., $\hat{\mathbf{t}}$ does not lie on the great circle defined by $\mathbf{t}$ and $\mathbf{s}$) and let us change $\hat{\mathbf{t}}$ such that $(\mathbf{t} \times \mathbf{s}) \cdot \hat{\mathbf{t}} = 0$. As $\hat{\mathbf{t}}$ changes, the area of type II on the sphere becomes an area of type IV and the area of type III becomes an area of type I. Thus, the negative range volume obtained consists only of range values above areas of type I.

Let us use the following notation. A_{III-I} denotes the area which changes from type III to type I and V_{III} and $V_{I(III)}$ are the volumes before and after change. Similarly, A_{II-IV} denotes the area which changes from type II to type IV and V_{II} and V_{IV} are the corresponding volumes.

The change of $\hat{\mathbf{t}}$ does not have any effect on the volumes above the areas that did not change in type, as can be seen from the constraint on $|\mathbf{R}|$ in Table 1. However, the change of $\hat{\mathbf{t}}$ causes a decrease in the volume above the areas which changed in type: Volume $V_{I(III)} < V_{II}$. Furthermore, as can be seen from equation (8), the normalization term is the same for points $\mathbf{r}_1(\varphi_{x_1}, \varphi_{y_1})$ and $\mathbf{r}_2(\varphi_{x_2}, \varphi_{y_2})$ symmetric with respect to the great circle $\mathbf{s} \cdot \mathbf{r} = 0$, because $\varphi_{y_1} = \varphi_{y_2}$ and $\varphi_{x_1} + \varphi_{x_2} = 2k\pi$ with $k \in \mathbb{N}$. Thus we encounter the same normalization factors in areas A_{III-I} and A_{II-IV}.

The volume of negative range values for any $\mathbf{s}$ is smallest for $(\mathbf{t} \times \mathbf{s}) \cdot \hat{\mathbf{t}} = 0$, independent of the range of values in which the scene lies. If we assume an upper bound $|\mathbf{R}_{\max}| \neq \infty$, or a lower bound $|\mathbf{R}_{\min}| \neq 0$, or both bounds on the scene in view, there exist points $\mathbf{r}$ in areas I and III above which there are no range values which contribute to the negative range volume. However as shown before, since the curves $C_{\min} = 0$, $C_{\max} = 0$, $(\boldsymbol{\omega}_\epsilon \cdot \mathbf{r})(\mathbf{s} \cdot \mathbf{r}) = 0$ and $(\mathbf{t} \times \mathbf{s}) \cdot \mathbf{r} = 0$ intersect at the same point, V_{II} must always be larger than $V_{I(III)}$.

For any $\mathbf{s}$ the smallest volume is obtained for $\mathbf{s}$, $\mathbf{t}$, and $\hat{\mathbf{t}}$ lying on a great circle. Therefore, in order to minimize the total negative range volume V_n, we must have $\mathbf{t} = \hat{\mathbf{t}}$.

Thus, in summary, we have shown that for any given rotational error $\boldsymbol{\omega}_\epsilon$ the negative range volume is smallest if the direction of the actual translation and the estimated translation coincide, that is, $\mathbf{t} = \hat{\mathbf{t}}$.

3.2 Fixed translational error

In this section we choose the following parameterization: The unit vectors $\mathbf{t}$ and $\hat{\mathbf{t}}$ lie in the yz plane, and $(\mathbf{t} \cdot \hat{\mathbf{t}}) \geq 0$; $\mathbf{s}$ lies in the xz plane with $s = (\sin\chi, 0, \cos\chi)$ and $\chi \in [0, \pi]$. As before, $\mathbf{n} = \frac{\mathbf{s} \times \mathbf{r}}{|\mathbf{s} \times \mathbf{r}|}$. The normalization term necessary for the uniformity of the orientations $\mathbf{n}$ in this parameterization can be obtained from equation (8) by substituting φ_y for φ_x and φ_z for φ_y, that is,

$$\left| \frac{\partial g(\chi)}{\partial(\chi)} \right| = \left| \frac{\sin\varphi_z}{\cos(\varphi_z)^2 \cos(\chi - \varphi_y)^2 - 1} \right|$$

where φ_z is the angle between $\mathbf{r}$ and the xz plane and φ_y is the angle between the projection of $\mathbf{r}$ on the xz plane and some fiducial direction in the xz plane.

We are given $\mathbf{t}$ and $\hat{\mathbf{t}}$, and we are interested in the direction of $\boldsymbol{\omega}_\epsilon$ minimizing the negative range volume. In analog to Section 3.1, we study the different areas on the sphere with unbounded and bounded intervals of estimated negative range values. Requiring

that $|\hat{\mathbf{R}}|$ be negative we obtain the inequality

$$|\hat{\mathbf{R}}| = |\mathbf{R}| \frac{(\hat{\mathbf{t}} \times \mathbf{s}) \cdot \mathbf{r}}{(\mathbf{t} \times \mathbf{s}) \cdot \mathbf{r} - |\mathbf{R}| ((\omega_\epsilon \cdot \mathbf{r})(\mathbf{s} \cdot \mathbf{r}) - (\omega_\epsilon \cdot \mathbf{r}))} < 0 \tag{11}$$

and thus the three curves $(\mathbf{t} \times \mathbf{s}) \cdot \mathbf{r} = 0$, $(\hat{\mathbf{t}} \times \mathbf{s}) \cdot \mathbf{r} = 0$ and $g = (\omega_\epsilon \cdot \mathbf{r})(\mathbf{s} \cdot \mathbf{r}) - (\omega_\epsilon \cdot \mathbf{s}) = 0$ separating areas I to IV.[1] The classification is analogous to the one in Table 1, except that the term $(\omega_\epsilon \cdot \mathbf{r})(\mathbf{s} \cdot \mathbf{r})$ must be replaced by the term $(\omega_\epsilon \cdot \mathbf{r})(\mathbf{s} \cdot \mathbf{r}) - (\omega_\epsilon \cdot \mathbf{s})$. Figure 8 provides a pictorial description of this classification for the case of $\mathbf{s}$ outside the interval $[\mathbf{t}, \hat{\mathbf{t}}]$, that is, $\text{sgn}(\mathbf{t} \times \mathbf{s}) = \text{sgn}(\hat{\mathbf{t}} \times \mathbf{s})$, and the case of $\mathbf{s}$ between $\mathbf{t}$ and $\hat{\mathbf{t}}$, that is, $\text{sgn}(\mathbf{t} \times \mathbf{s}) = -\text{sgn}(\hat{\mathbf{t}} \times \mathbf{s})$. If $\text{sgn}(\mathbf{t} \times \mathbf{s}) = \text{sgn}(\hat{\mathbf{t}} \times \mathbf{s})$ the sphere is separated into areas I and IV, each covering an area the size of a hemisphere; otherwise, the sphere is separated into areas II and III, again each the size of a hemisphere.

Again, if we consider a lower bound $|\mathbf{R}_{\min}|$ and an upper bound $|\mathbf{R}_{\max}|$ for the scene in view, we obtain the two curves $C_{\min} = (\mathbf{t} \times \mathbf{s}) \cdot \mathbf{r} - |\mathbf{R}_{\min}| ((\omega_\epsilon \cdot \mathbf{r})(\mathbf{s} \cdot \mathbf{r}) - \omega_\epsilon \cdot \mathbf{s})) = 0$ and $C_{\max} = (\mathbf{t} \times \mathbf{s}) \cdot \mathbf{r} - |\mathbf{R}_{\max}| ((\omega_\epsilon \cdot \mathbf{r})(\mathbf{s} \cdot \mathbf{r}) - \omega_\epsilon \cdot \mathbf{s})) = 0$, as shown in Figure 8. $C_{\min} = 0$ and $C_{\max} = 0$ separate the points $\mathbf{r}$ in areas I and III into those with no volume of negative range values, those with a volume bounded by a value different from $|\mathbf{R}_{\min}|$ and $|\mathbf{R}_{\max}|$, and those with a volume ranging from $|\mathbf{R}_{\min}|$ to $|\mathbf{R}_{\max}|$.

The proof is given in three parts. We decompose ω_ϵ into a component ω_{par} which lies in the xz plane and a component $\omega_{\text{perp}} = \lambda(0, -1, 0)$ parallel to the y axis: $\omega_\epsilon = \omega_{\text{par}} + \omega_{\text{perp}}$. First, we show that if $\omega_{\text{par}} = 0$, the smallest negative depth volume is obtained for $\omega_{\text{perp}} = 0$. Second, we show that if $\omega_{\text{perp}} = 0$, a vector $\omega_{\text{par}} \neq 0$ with $(\mathbf{t} \times \omega_{\text{par}}) = -(\hat{\mathbf{t}} \times \omega_{\text{par}})$ is obtained, which we call ω_{par_0}, that provides the smallest negative range volume. Third, we prove that ω_ϵ, in order to minimize the negative range volume, must satisfy the constraint $(\omega_\epsilon \cdot \mathbf{t}) = (\omega_\epsilon \cdot \hat{\mathbf{t}})$. However, if we change the direction of ω_ϵ, which amounts to $\frac{\omega_{\text{par}} + \omega_{\text{perp}}}{|\omega_{\text{par}} + \omega_{\text{perp}}|}$ with $\omega_{\text{perp}} = \lambda(0, -1, 0)$, by continuously increasing $\lambda \geq 0$, the negative depth volume increases monotonically, and thus the smallest negative depth volume is obtained for ω_{par_0}. The details of the proofs will now be given.

Part 1 (ω_{perp} minimizing the negative range volume)

Let $\omega_{\text{par}} = 0$; then $g = (\omega_{\text{perp}} \cdot \mathbf{r})(\mathbf{s} \cdot \mathbf{r}) = 0$, and the curves $C_i = 0$ for $i = \{\min, \max\}$ become

$$C_i = (\mathbf{t} \times \mathbf{s}) \cdot \mathbf{r} - |\mathbf{R}_i| (\omega_{\text{perp}} \cdot \mathbf{r})(\mathbf{s} \cdot \mathbf{r}) = 0$$

Since $(\mathbf{t} \times \mathbf{s}) \times \omega_{\text{perp}} = 0$,

$$C_i = (\omega_{\text{perp}} \cdot \mathbf{r}) \left(\frac{\sin \angle(\mathbf{t}, \mathbf{s}) |\mathbf{t}|}{|\omega_{\text{perp}}| \, |\mathbf{R}_i|} - (\mathbf{s} \cdot \mathbf{r}) \right) = 0$$

where $\sin \angle(\mathbf{t}, \mathbf{s})$ denotes the sine of the angle between vectors $\mathbf{t}$ and $\mathbf{s}$. Curve $C_i = 0$ thus consists of the great circle $\omega_{\text{perp}} \cdot \mathbf{r} = 0$ and the circle $\frac{\sin \angle(\mathbf{t},\mathbf{s}) |\mathbf{t}|}{|\omega_{\text{perp}}| |\mathbf{R}_i|} - (\mathbf{s} \cdot \mathbf{r}) = 0$ parallel

[1] Curve $g = 0$ is of the same form as the zero-motion contour defined in [13], which got its name from the fact that it defines the locus of points for which a flow field due to rigid motion with translation $\mathbf{s}$ and rotation ω can take on the value zero.

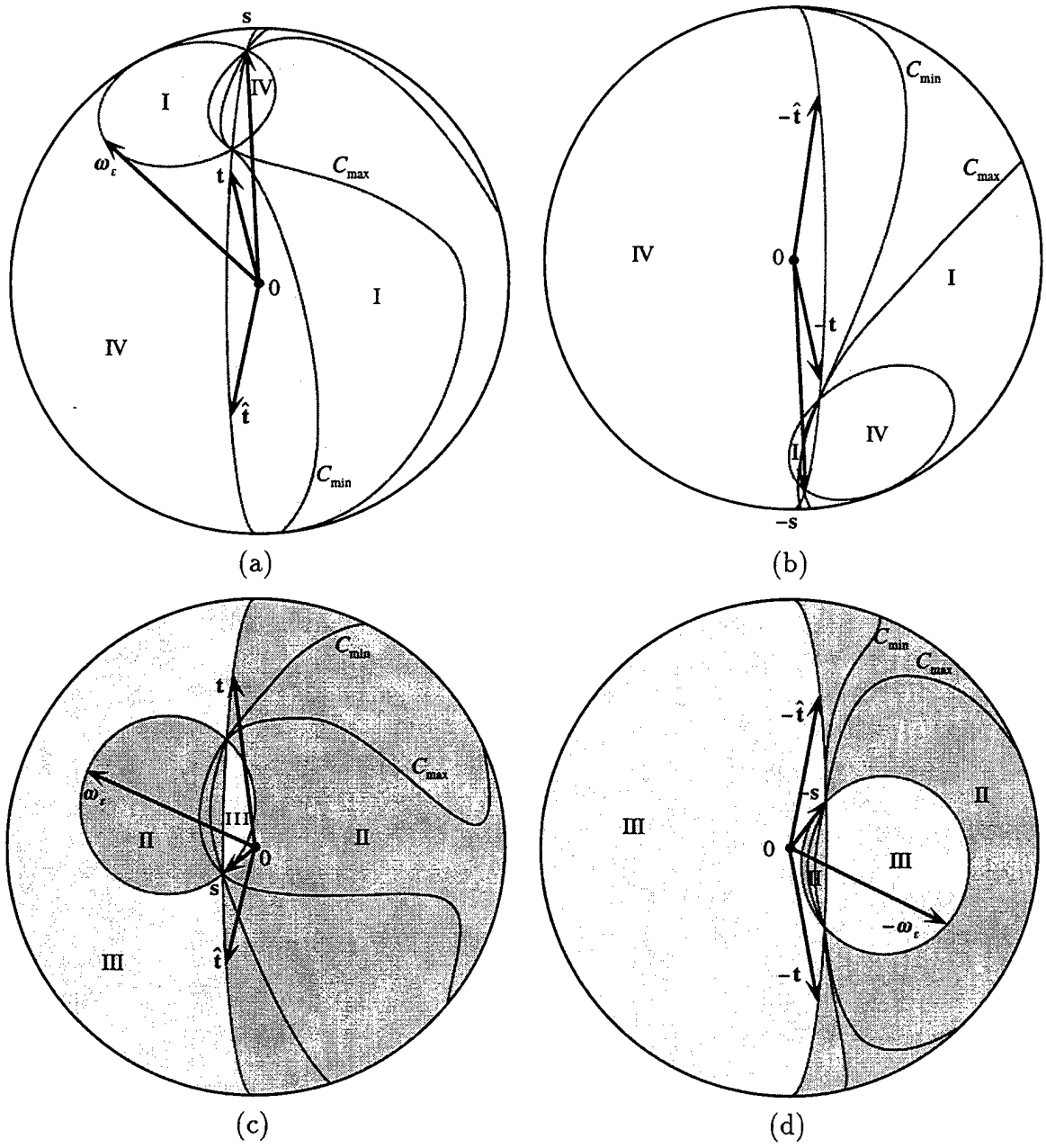


Figure 8: Classification of image points for general $\boldsymbol{\omega}_\epsilon$. (a), (b): $\text{sgn}(\mathbf{t}\times\mathbf{s})\cdot\mathbf{r} = \text{sgn}(\hat{\mathbf{t}}\times\mathbf{s})\cdot\mathbf{r}$. (c), (d): $\text{sgn}(\mathbf{t}\times\mathbf{s})\cdot\mathbf{r} = -\text{sgn}(\hat{\mathbf{t}}\times\mathbf{s})\cdot\mathbf{r}$. If $\text{sgn}(\mathbf{t}\times\mathbf{s}) = \text{sgn}(\hat{\mathbf{t}}\times\mathbf{s})$ the negative estimated range values are in area I above the area defined by curves $C_{\max} = 0$ and $(\mathbf{t}\times\mathbf{s})\cdot\mathbf{r} = 0$. If $\text{sgn}(\mathbf{t}\times\mathbf{s}) = -\text{sgn}(\hat{\mathbf{t}}\times\mathbf{s})$ the negative estimated range values are above area II and in area III above the area defined by curves $C_{\min} = 0$ and $(\boldsymbol{\omega}_\epsilon\cdot\mathbf{r})(\mathbf{s}\cdot\mathbf{r}) - (\boldsymbol{\omega}_\epsilon\cdot\mathbf{s}) = 0$.

to the great circle $(\mathbf{s} \cdot \mathbf{r}) = 0$. If $\frac{\sin \angle(\mathbf{t},\mathbf{s})|\mathbf{t}|}{|\boldsymbol{\omega}_{\text{perp}}||\mathbf{R}_i|} > 1$ this circle disappears. Figure 9 provides a pictorial description of the areas I–IV and the curves $C_i = 0$.

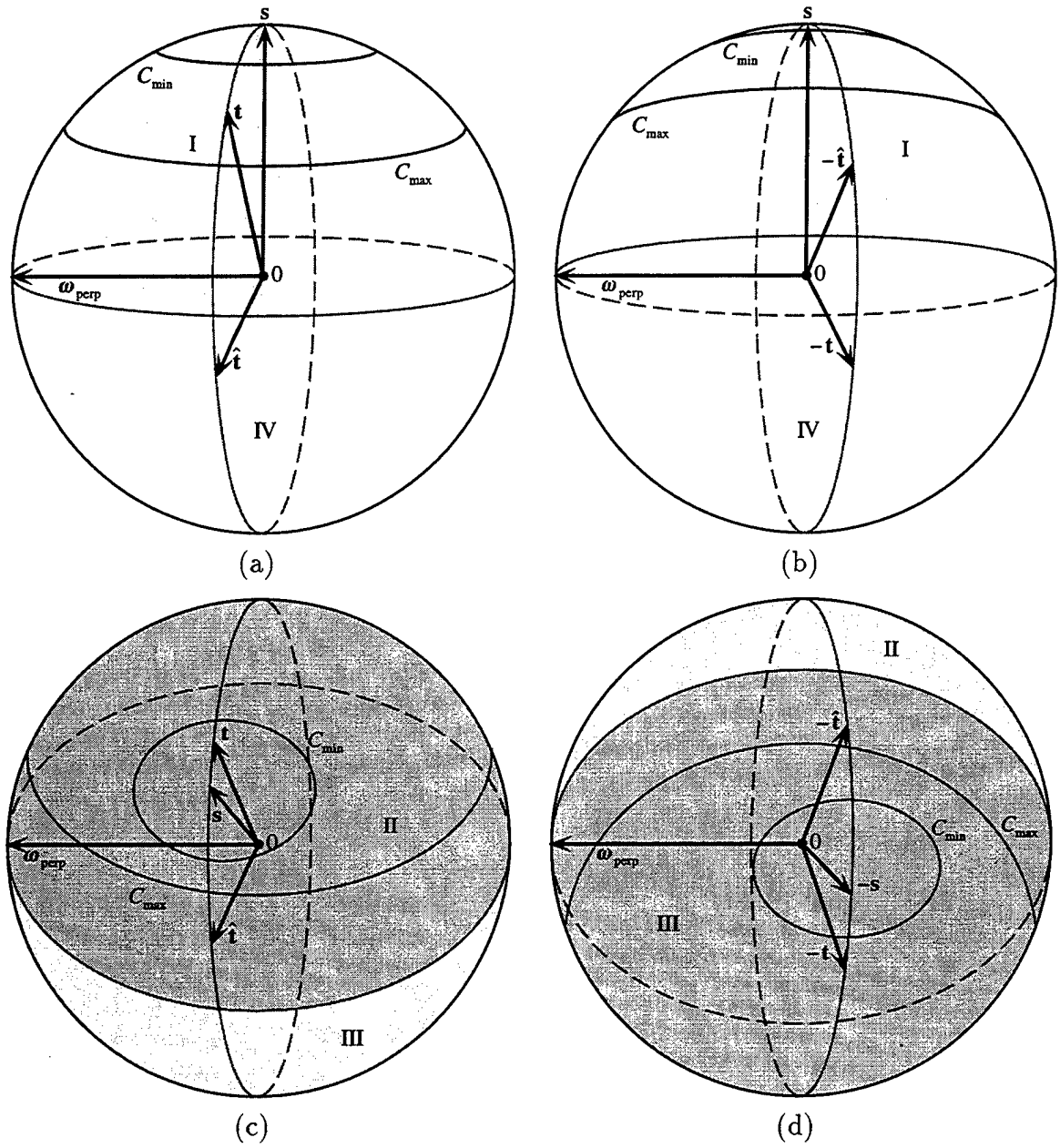


Figure 9: Classification of image points for $\boldsymbol{\omega}_\epsilon = \boldsymbol{\omega}_{\text{perp}}$.

Let us consider two flow directions defined by vectors $\mathbf{s}_1$ and $\mathbf{s}_2$ that are symmetric with regard to $\mathbf{t}$, that is $(\mathbf{s}_1 \times \mathbf{t}) = -(\mathbf{s}_2 \times \mathbf{t})$, and let $\mathbf{s}_1$ be between $\mathbf{t}$ and $\hat{\mathbf{t}}$. For every point $\mathbf{r}_1$ in area III defined by $\mathbf{s}_1$ there exists a point $\mathbf{r}_2$ in area I defined by $\mathbf{s}_2$ with the same normalization factor such that the negative estimated ranges above $\mathbf{r}_1$ and $\mathbf{r}_2$ add up to $|\mathbf{R}_{\max}| - |\mathbf{R}_{\min}|$. Thus, the volume of negative estimated range obtained from $\mathbf{s}_1$ and $\mathbf{s}_2$, denoted by $V(\mathbf{s}_1)$ and $V(\mathbf{s}_2)$ respectively, amounts to the area of the sphere A_{sp} times

$|\mathbf{R}_{\max}| - |\mathbf{R}_{\min}|$, that is, $V(\mathbf{s}_1) + V(\mathbf{s}_2) = A_{\text{sp}}(|\mathbf{R}_{\max}| - |\mathbf{R}_{\min}|)$ (area II of $\mathbf{s}_1$ contributes a hemisphere; area III of $\mathbf{s}_1$ and area I of $\mathbf{s}_2$ together contribute a hemisphere). Let us decompose the negative range volume and provide the following definition:

$$V_n = V_A + V_B \quad \text{and} \quad V_A = V_{A'} + V_{A''}$$

V_A corresponds to the volume obtained from all $\mathbf{s}$ with $|\mathbf{t} \cdot \mathbf{s}| \geq |\mathbf{t} \cdot \hat{\mathbf{t}}|$. $V_{A'}$ corresponds to those $\mathbf{s}$ with $\text{sgn}(\mathbf{t} \cdot \mathbf{s}) = -\text{sgn}(\hat{\mathbf{t}} \cdot \mathbf{s})$ (that is, all $\mathbf{s}$ between $\mathbf{t}$ and $\hat{\mathbf{t}}$) and $V_{A''}$ corresponds to those $\mathbf{s}$ with $\text{sgn}(\mathbf{t} \cdot \mathbf{s}) = \text{sgn}(\hat{\mathbf{t}} \cdot \mathbf{s})$ (that is, the set of $\mathbf{s}$ symmetric in $\mathbf{t}$ to the set of $\mathbf{s}$ in $V_{A'}$). V_B corresponds to the volume from all the remaining $\mathbf{s}$'s. $V_{A'}$ consists of range values above areas of type III and II, and $V_{A''}$ and V_B consist of range values above areas of type I only.

$V_A = V_{A'} + V_{A''} = A_{\text{sp}}(|\mathbf{R}_{\max}| - |\mathbf{R}_{\min}|)\angle(\mathbf{t}, \hat{\mathbf{t}})$ and $V_B \geq 0$. $V_B = 0$ if for all $\mathbf{s}$ contributing to volume V_B we have $\frac{\sin \angle(\mathbf{t},\mathbf{s})|\mathbf{t}|}{|\mathbf{R}_{\max}||\boldsymbol{\omega}_{\text{perp}}|} \geq 1$, that is, if

$$\frac{\sin \angle(\mathbf{t}, \hat{\mathbf{t}})|\mathbf{t}|}{|\mathbf{R}_{\max}|\,|\boldsymbol{\omega}_{\text{perp}}|} \geq 1$$

Thus, in summary, if $\boldsymbol{\omega}_{\text{perp}} = 0$, the minimum negative range volume V_n is equal to $A_{\text{sp}}(|\mathbf{R}_{\max}| - |\mathbf{R}_{\min}|)\angle(\mathbf{t}, \hat{\mathbf{t}})$ and is obtained for all $\boldsymbol{\omega}_{\text{perp}}$ with $|\boldsymbol{\omega}_{\text{perp}}| \leq \frac{\sin \angle(\mathbf{t},\hat{\mathbf{t}})|\mathbf{t}|}{|\mathbf{R}_{\max}|}$. If $|\mathbf{R}_{\max}| = \infty$ or $(\mathbf{t} \cdot \hat{\mathbf{t}}) = 1$ there is only one solution, $\boldsymbol{\omega}_{\text{perp}} = 0$.

Part 2 ($\boldsymbol{\omega}_{\text{perp}}$ minimizing the negative range volume)

If $\boldsymbol{\omega}_{\text{perp}} = 0$, curve $g = (\boldsymbol{\omega}_{\text{par}} \cdot \mathbf{r})(\mathbf{s} \cdot \mathbf{r}) - (\boldsymbol{\omega}_{\text{par}} \cdot \mathbf{s}) = 0$ is symmetric with respect to the xz plane and curves $C_i = 0$ for $i = \{\min, \max\}$ become (see Figure 10)

$$C_i = (\mathbf{t} \times \mathbf{s}) \cdot \mathbf{r} - |\mathbf{R}_i|\,((\boldsymbol{\omega}_{\text{par}} \cdot \mathbf{r})(\mathbf{s} \cdot \mathbf{r}) - (\boldsymbol{\omega}_{\text{par}} \cdot \mathbf{s})) = 0$$

Let us fix $\mathbf{s}$ and $|\boldsymbol{\omega}_{\text{par}}|$ and let us vary the direction $\frac{\boldsymbol{\omega}_{\text{par}}}{|\boldsymbol{\omega}_{\text{par}}|}$. As $\angle(\mathbf{s}, \boldsymbol{\omega}_{\text{par}})$ increases, the area between $C_{\min} = 0$ and $C_{\max} = 0$ multiplied by the normalization factor, and the area between $C_{\min} = 0$ and $(\mathbf{t} \times \mathbf{s}) \cdot \mathbf{r} = 0$ multiplied by the normalization factor, decrease. This can be verified by numerical integration. It can also be understood from the following observation: Referring to Figure 11, we see that if $(\mathbf{s} \cdot \boldsymbol{\omega}_{\text{par}}) \neq 0$, an increase in $\angle(\mathbf{s}, \boldsymbol{\omega}_{\text{par}})$ causes an increase in the size of curve $g_i = 0$. Therefore, the area between $C_i = 0$ and $(\mathbf{t} \times \mathbf{s}) \cdot \mathbf{r} = 0$ in the left hemisphere increases, but the area between $C_i = 0$ and $(\mathbf{t} \times \mathbf{s}) \cdot \mathbf{r}$ in the right hemisphere decreases by a larger amount, since the area inside curve $g = 0$ is smaller than the area inside curve $(\mathbf{t} \times \mathbf{s}) \cdot \mathbf{r} = 0$. Furthermore, the normalization factors in the area added in the left hemisphere are smaller than the normalization factors in the area lost in the right hemisphere. Therefore, if $\text{sgn}(\mathbf{t} \times \mathbf{s}) = \text{sgn}(\hat{\mathbf{t}} \times \mathbf{s})$, the negative range volume above area I decreases, and if $\text{sgn}(\mathbf{t} \times \mathbf{s}) = -\text{sgn}(\hat{\mathbf{t}} \times \mathbf{s})$ the negative range volume above area III increases as $\angle(\mathbf{s}, \boldsymbol{\omega}_{\text{par}})$ increases.

The negative range volume $V(\mathbf{s})$ for each $\mathbf{s}$ can be decomposed into a component $V_1(\mathbf{s})$ dependent on $(\mathbf{t} \times \mathbf{s}) \cdot \mathbf{r}$ and a component $V_2(\mathbf{s})$ dependent on $|\mathbf{R}_i|\,((\boldsymbol{\omega}_{\text{par}} \cdot \mathbf{r})(\mathbf{s} \cdot \mathbf{r}) - (\boldsymbol{\omega}_{\text{par}} \cdot \mathbf{s}))$. The overall negative range volume V_n is obtained by integrating $V(\mathbf{s})$ over all $\mathbf{s}$, that is, $V_n = \int_s V(\mathbf{s})\,d\mathbf{s} = \int_s V_1(\mathbf{s})\,d\mathbf{s} + \int_s V_2(\mathbf{s})\,d\mathbf{s}$. Since $V_1(\mathbf{s})$ does not depend on $\boldsymbol{\omega}_{\text{par}}$, V_n will

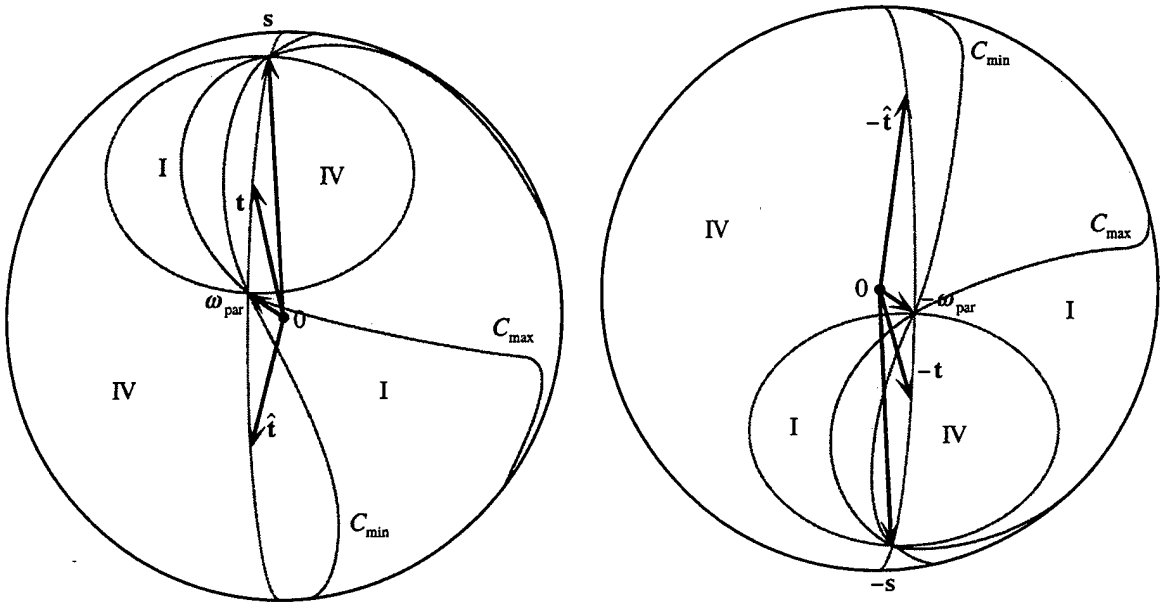


Figure 10: Classification of image points for $\omega_\epsilon = \omega_{\text{par}}$.

attain its minimum at the minimum of $\int_s V_2(\mathbf{s})\, d\mathbf{s}$, which is achieved when ω_{par} is between $\mathbf{t}$ and $\hat{\mathbf{t}}$ at an equal distance from both, that is, if $(\mathbf{t} \times \omega_{\text{par}}) = -(\hat{\mathbf{t}} \times \omega_{\text{par}})$.

Let us denote by ω_{par_0} the ω_{par} minimizing the negative range volume V_n. We have found the direction of ω_{par_0}; it remains to be shown that $|\omega_{\text{par}_0}| \neq 0$.

For $\omega_{\text{par}_0} = 0$, $V_n = A_{\text{sp}}(|\mathbf{R}_{\max}| - |\mathbf{R}_{\min}|)\angle(\mathbf{t}\hat{\mathbf{t}})$. $V_n = V_A + V_B = V_{A'} + V_{A''} + V_B$. Since for any $\mathbf{s}_1$ with $\text{sgn}(\mathbf{t} \times \mathbf{s}_1) = -\text{sgn}(\hat{\mathbf{t}} \times \mathbf{s}_1)$, and any $\mathbf{s}_2$ with $\text{sgn}(\mathbf{t} \times \mathbf{s}_2) = \text{sgn}(\hat{\mathbf{t}} \times \mathbf{s}_2)$, $V(\mathbf{s}_1) + V(\mathbf{s}_2) < A_{\text{sp}}(|\mathbf{R}_{\max}| - |\mathbf{R}_{\min}|)$, it follows that volume $V_A = V_{A'} + V_{A''} < A_{\text{sp}}(|\mathbf{R}_{\max}| - |\mathbf{R}_{\min}|)(\angle(\mathbf{t}\hat{\mathbf{t}}))$. Volume $V_B \geq 0$. If $(\mathbf{t} \cdot \hat{\mathbf{t}}) = 0$ (that is, $\angle(\mathbf{t}\hat{\mathbf{t}}) = \pi/2$), volume $V_B = 0$, and thus there exists a $\omega_{\text{par}_0} \neq 0$ minimizing the negative range volume. Since due to the symmetry on the sphere, $|\omega_{\text{par}_0}|$ must change monotonically as $(\mathbf{t} \cdot \hat{\mathbf{t}})$ increases, we conclude that $\omega_{\text{par}_0} \neq 0$ for all $(\mathbf{t} \cdot \hat{\mathbf{t}}) > 0$.

Part 3 ($\omega_\epsilon = \omega_{\text{par}} + \omega_{\text{perp}}$)

Let us consider $\omega_\epsilon = \omega_{\text{par}} + \omega_{\text{perp}}$, with ω_{par} a component in the xz plane and $\omega_{\text{perp}} = \lambda(0, -1, 0)$ parallel to the y axis. Therefore,

$$g = (\omega_{\text{perp}} \cdot \mathbf{r})(\mathbf{s} \cdot \mathbf{r}) + (\omega_{\text{par}} \cdot \mathbf{r})(\mathbf{s} \cdot \mathbf{r}) - (\omega_{\text{par}} \cdot \mathbf{s}) = 0$$

and

$$C_i = (\mathbf{t} \times \mathbf{s}) \cdot \mathbf{r} - |\mathbf{R}_i| \left((\omega_{\text{perp}} \cdot \mathbf{r})(\mathbf{s} \cdot \mathbf{r}) + (\omega_{\text{par}} \cdot \mathbf{r})(\mathbf{s} \cdot \mathbf{r}) - (\omega_{\text{par}} \cdot \mathbf{s}) \right) = 0$$

Let us fix $|\omega_\epsilon| = 1$ and change ω_ϵ by increasing $|\omega_{\text{perp}}|$ and thus increasing λ. If $(\omega_{\text{par}} \cdot \mathbf{s}) > 0$ an increase in λ causes a decrease in the area between $C_{\min} = 0$ and $C_{\max} = 0$ times the normalization factor, and a decrease in the area between $C_{\min} = 0$

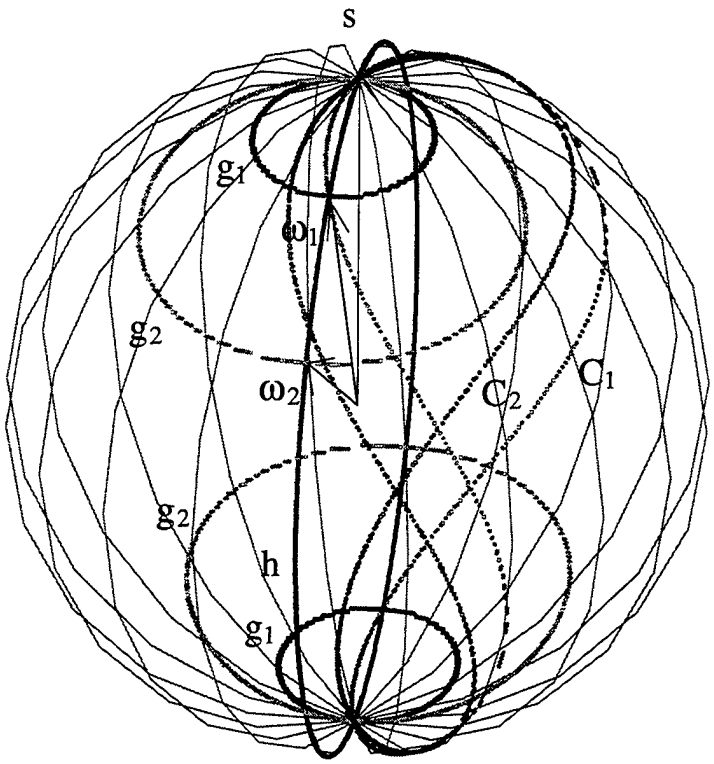


Figure 11: As ω_{par} changes from ω_1 to ω_2 and $\angle(\omega_{\text{par}}, \mathbf{s})$ increases, curve $g = 0$ changes from g_1 to g_2 and curve $C_i = 0$ changes from C_1 to C_2. In areas of type I the volume between $C_i = 0$ and $h = (\mathbf{t} \times \mathbf{s}) \cdot \mathbf{r} = 0$ decreases and in areas of type III the volume between $C_i = 0$ and $g = 0$ increases.

and $(\mathbf{t} \times \mathbf{s}) \cdot \mathbf{r} = 0$ times the normalization factor. Again, this can be verified either by numerical integration or through the following observation in reference to Figure 12: An increase in λ causes an increase in the size of $g = 0$. Since $g = 0$ is not symmetric about the xz plane, the area between $C_i = 0$ and $(\mathbf{t} \times \mathbf{s}) \cdot \mathbf{r} = 0$ in the left hemisphere times the normalization factor increases, but the area between $C_i = 0$ and $(\mathbf{t} \times \mathbf{s}) \cdot \mathbf{r} = 0$ in the right hemisphere times the normalization factor decreases by a larger amount.

Thus, an increase in λ has the effect that for all $\mathbf{s}$ with $\mathrm{sgn}(\mathbf{t} \times \mathbf{s}) = -\mathrm{sgn}(\hat{\mathbf{t}} \times \mathbf{s})$ the negative range volume above area III increases, and for all $\mathbf{s}$ with $\mathrm{sgn}(\mathbf{t} \times \mathbf{s}) = \mathrm{sgn}(\hat{\mathbf{t}} \times \mathbf{s})$ the volume above area I decreases. It can further be observed that the larger $(\boldsymbol{\omega}_{\mathrm{par}} \cdot \mathbf{s})$, the larger the decrease in the volume above area I (or the larger the increase in the volume above area III), with a peak for $(\boldsymbol{\omega}_{\mathrm{par}} \cdot \mathbf{s}) = 1$.

Using the same argument as before, that the negative range volume V_n can be decomposed into a component due to $(\mathbf{t} \times \mathbf{s}) \cdot \mathbf{r}$ only and a component due to $|\mathbf{R}_i|\,(\boldsymbol{\omega}_\epsilon \cdot \mathbf{r})(\mathbf{s} \cdot \mathbf{r}) - (\boldsymbol{\omega}_\epsilon \cdot \mathbf{s})$ only, we find that the volume above areas of type III must be as small as possible, and thus $(\boldsymbol{\omega}_\epsilon \cdot \mathbf{t}) = (\boldsymbol{\omega}_\epsilon \cdot \hat{\mathbf{t}})$. Therefore $\boldsymbol{\omega}_\epsilon$ can only be of the form

$$\boldsymbol{\omega}_\epsilon = \lambda \left(\boldsymbol{\omega}_{\mathrm{par}_0} + \boldsymbol{\omega}_{\mathrm{perp}} \right)$$

The smallest negative range volume is smaller for $\boldsymbol{\omega}_{\mathrm{par}_0}$ than for any $\boldsymbol{\omega}_{\mathrm{perp}}$.

Again we decompose V_n: $V_n = V_A + V_B = V_{A'} + V_{A''} + V_B$. If $(\mathbf{t} \cdot \hat{\mathbf{t}}) = 0$, $V_B = 0$, and since an increase in λ causes an increase in $V_{A'}$ which is larger than the decrease in $V_{A''}$, the negative range volume must increase monotonically and $|\boldsymbol{\omega}_\epsilon|$ must decrease monotonically. Therefore, if $(\mathbf{t} \cdot \hat{\mathbf{t}}) > 0$ and $(\mathbf{t} \cdot \hat{\mathbf{t}}) \neq 1$ the smallest negative range volume must also increase monotonically and the $|\boldsymbol{\omega}_\epsilon|$ minimizing the volume must decrease monotonically.

Thus in summary we have shown that for a given $\mathbf{t}$ and $\hat{\mathbf{t}}$, the rotational error $\boldsymbol{\omega}_\epsilon$ which minimizes the negative range volume is $\boldsymbol{\omega}_{\mathrm{par}_0} \neq 0$. The direction of $\boldsymbol{\omega}_{\mathrm{par}_0}$ is such that $(\boldsymbol{\omega}_{\mathrm{par}_0} \times \mathbf{t}) = -(\boldsymbol{\omega}_{\mathrm{par}_0} \times \hat{\mathbf{t}})$.

4 The Planar Case

Let us express equation (4) in the more common component notation: $\dot{\mathbf{r}} = (\dot{r}_1, \dot{r}_2, \dot{r}_3)$. $\dot{r}_3$ is zero. If we denote $\dot{r}_1$ by u and $\dot{r}_2$ by v and express the coordinates of the focus of expansion as $(x_0, y_0) = (\frac{Uf}{W}, \frac{Vf}{W})$ we obtain the well-known equations

$$\begin{aligned} u &= \tfrac{u_{\mathrm{tr}}}{Z} + u_{\mathrm{rot}} = (x - x_0)\tfrac{W}{Z} + \alpha\tfrac{xy}{f} - \beta\left(\tfrac{x^2}{f} + f\right) + \gamma y \\ v &= \tfrac{v_{\mathrm{tr}}}{Z} + v_{\mathrm{rot}} = (y - y_0)\tfrac{W}{Z} + \alpha\left(\tfrac{y^2}{f} + f\right) - \beta\tfrac{xy}{f} - \gamma x \end{aligned}$$

Since, due to the scaling ambiguity, only the direction of translation can possibly be obtained, we set $W = 1$ and obtain from equation (7)

$$\hat{Z} = Z \left(\frac{(x - \hat{x}_0)\, n_x + (y - \hat{y}_0)\, n_y}{\begin{array}{c}(x - x_0)\, n_x + (y - y_0)\, n_y + Z\Big(\left(\alpha_\epsilon \frac{xy}{f} - \beta_\epsilon \left(\frac{x^2}{f} + f\right) + \gamma_\epsilon y\right) n_x \\ + \left(\alpha_\epsilon \left(\frac{y^2}{f} + f\right) - \beta_\epsilon \frac{xy}{f} - \gamma_\epsilon x\right) n_y \Big)\end{array}} \right) \tag{12}$$

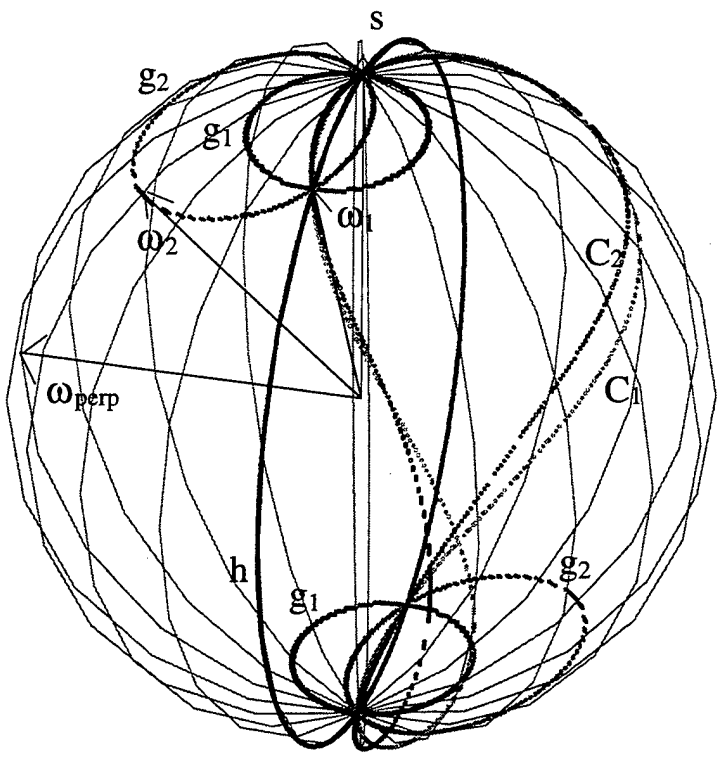


Figure 12: As ω_ϵ changes from $\omega_1 = \frac{\omega_{\text{par}}}{|\omega_{\text{par}}|}$ to $\omega_2 = \frac{\omega_{\text{par}}+\lambda\omega_{\text{perp}}}{|\omega_{\text{par}}+\lambda\omega_{\text{perp}}|}$, curve $g = 0$ changes from g_1 to g_2 and curve $C_i = 0$ changes from C_1 to C_2, the volume of negative range between $C_i = 0$ and $h = (\mathbf{t} \times \mathbf{s}) \cdot \mathbf{r} = 0$ decreases above area I, and the volume of negative range values between $C_i = 0$ and $g = 0$ increases above area III.

where n_x and n_y denote the components of $\mathbf{n}$ in the x and y directions.

In the following analysis, we assume that the FOE and the estimated FOE are within the image. We do not study any particular image shape, and we ignore the exact effects resulting from volumes of negative depth in different directions being outside the field of view. Such effects might introduce biases, but this is of little practical interest. Any implementation of algorithms based on the constraint of negative depth must consider the position of the FOE on the image, and cannot just be based on blindly counting the negative values. We also perform some simplification: For a limited field of view, the terms quadratic in the image coordinates, which appear in the rotational components, are small with respect to the linear and constant terms, and we therefore drop them.

The 0 distortion surface thus becomes

$$(x - \hat{x}_0)\, n_x + (y - \hat{y}_0)\, n_y = 0 \tag{13}$$

and the $-\infty$ distortion surface takes the form

$$(x - x_0)n_x + (y - y_0)n_y + Z\left((-\beta_\epsilon f + \gamma_\epsilon y)\, n_x + (\alpha_\epsilon f - \gamma_\epsilon x)\, n_y\right) = 0 \tag{14}$$

The flow directions (n_x, n_y) can alternatively be written as $(\cos\psi, \sin\psi)$, with $\psi \in [0, \pi]$ denoting the angle between $[n_x, n_y]^T$ and the x axis.

To simplify the visualization of the volumes of negative depth in different directions, we perform the following coordinate transformation to align the flow direction with the x axis: for every ψ we rotate the coordinate system by angle ψ, to obtain the new coordinates

$$[x', y']^T = R[x, y]^T, [x_0', y_0']^T = R[x_0, y_0]^T$$
$$[\hat{x}_0', \hat{y}_0']^T = R[x, y]^T, [\alpha_\epsilon', \beta_\epsilon']^T = R[\alpha_\epsilon, \beta_\epsilon]^T$$

where $R = \begin{bmatrix} \cos\psi & \sin\psi \\ -\sin\psi & \cos\psi \end{bmatrix}$.

Equations (13) and (14) thus become

$$(x' - \hat{x}_0') = 0$$
$$\text{and} \quad (x' - x_0') + Z\left(-\beta_\epsilon' f + \gamma_\epsilon y\right) = 0$$

In the following proof we first consider the case of $\gamma_\epsilon = 0$ and we then study the general case.

Part 1 ($\gamma_\epsilon = 0$)

If $\gamma_\epsilon = 0$, the volume of negative depth values for every direction ψ lies between the surfaces

$$(x' - \hat{x}_0') = 0 \quad \text{and} \quad (x' - x_0') - \beta_\epsilon' f Z = 0$$

Equation $(x' - \hat{x}_0') = 0$ describes a plane parallel to the $y'Z$ plane at distance $\hat{x}_0'$ from the origin, and equation $(x' - x_0') - \beta_\epsilon' f Z = 0$ describes a plane parallel to the y' axis of slope $\frac{1}{\beta_\epsilon' f}$, which intersects the $x'y'$ plane in the x' coordinate x_0'. Thus we obtain a wedge-shaped volume parallel to the y' axis. Figure 13 illustrates the volume through a slice parallel to the $x'Z$ plane and Figure 14 gives an illustration of this volume in space.

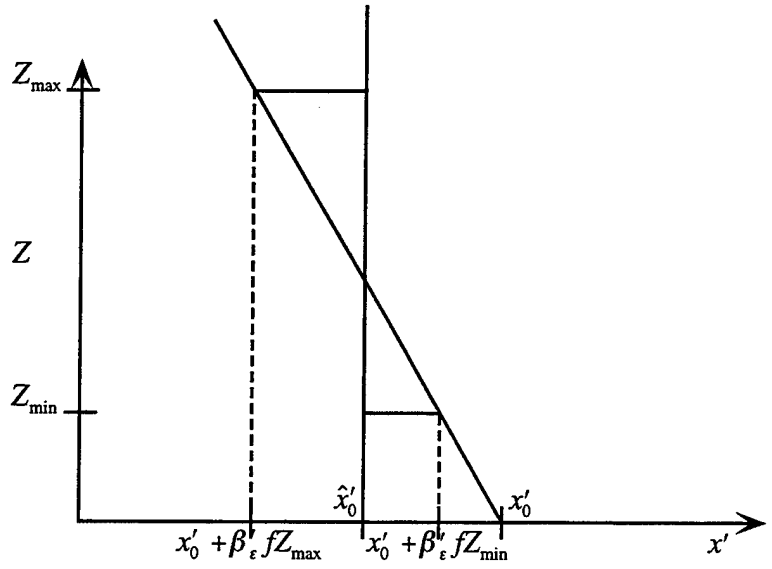


Figure 13: Slice parallel to the $x'Z$ plane through the volume of negative estimated depth for a single direction.

The scene in view extends between the depth values $Z_{\min}$ and $Z_{\max}$. The $-\infty$ distortion surface intersects the planes $Z = Z_{\max}$ and $Z = Z_{\min}$ in the x' coordinates ${x_0}' + {\beta_\epsilon}' f Z_{\max}$ and ${x_0}' + {\beta_\epsilon}' f Z_{\min}$. As can be seen from Figure 13, whether we are given the rotational error ${\beta_\epsilon}'$ or the translations ${x_0}'$ and $\hat{x}_0'$, and thus the translational error ${x_{0_\epsilon}}' = \cos\psi x_{0_\epsilon} + \sin\psi y_{0_\epsilon} = {x_0}' - \hat{x}_0'$, the minimum negative volume is obtained if $\hat{x}_0' = {x_0}' + {\beta_\epsilon}' f \frac{(Z_{\max}+Z_{\min})}{2}$. In other words, the 0 distortion surface has to intersect the $-\infty$ distortion surface in the middle of the depth interval in the plane $Z = \frac{Z_{\max}+Z_{\min}}{2}$.

Thus, for the direction defined by any angle ψ, the smallest volume of negative depth estimates is obtained if the rotational and translational errors are related as follows:

$$ {\beta_\epsilon}' = \frac{-2{x_{0_\epsilon}}'}{f\,(Z_{\max} + Z_{\min})} $$

Since ${\beta_\epsilon}' = \cos\psi\beta_\epsilon - \sin\psi\alpha_\epsilon$ and $x_{0_\epsilon} = \cos\psi x_{0_\epsilon} + \sin\psi y_{0_\epsilon}$, the volume is minimized for every direction if $\frac{x_{0_\epsilon}}{y_{0_\epsilon}} = -\frac{\beta_\epsilon}{\alpha_\epsilon}$. In other words, the rotational error $(\alpha_\epsilon, \beta_\epsilon)$ and the translational error $(x_{0_\epsilon}, y_{0_\epsilon})$ have to be perpendicular to each other.

Part 2 ($\gamma_\epsilon \neq 0$)

If $\gamma_\epsilon \neq 0$, the $-\infty$ distortion surface becomes

$$ (x' - \hat{x}_0') + Z\,(-{\beta_\epsilon}' f + \gamma_\epsilon y') = 0 $$

This surface can be most easily understood by slicing it with planes parallel to the $x'y'$ plane. At every depth value Z, we obtain a line of slope $\frac{-1}{\gamma_\epsilon Z}$ which intersects the x' axis in $x' = {x_0}' + {\beta_\epsilon}' f Z$ (see Figure 15). For any given Z the slopes of the lines in

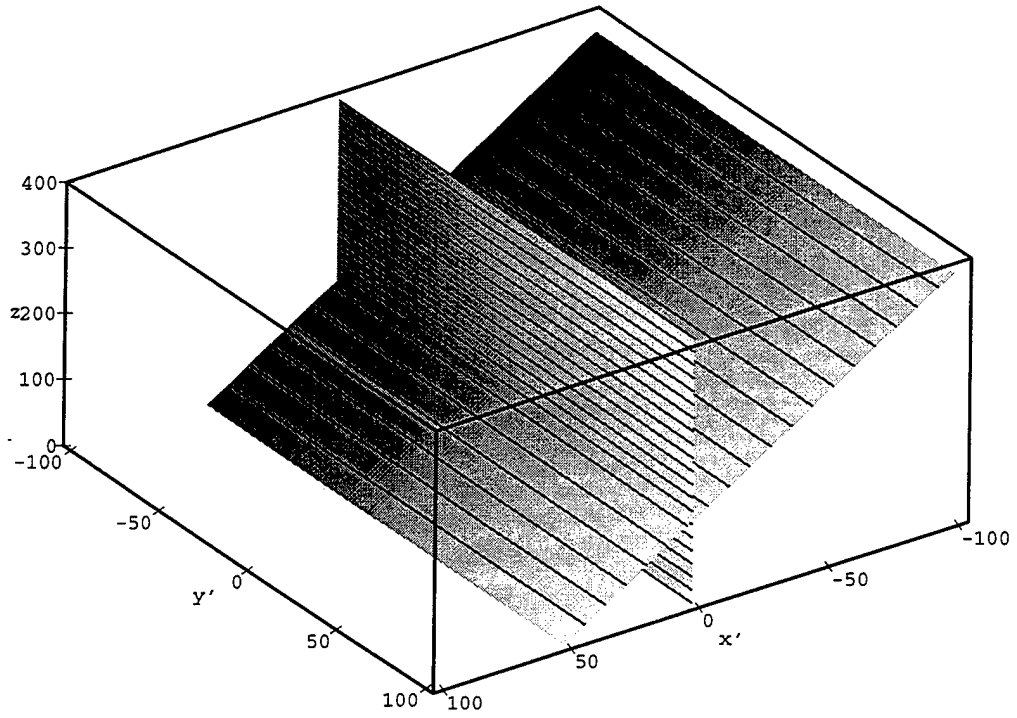


Figure 14: $\gamma_\epsilon = 0$: The volume of negative depth values for a single direction between the 0 and $-\infty$ distortion surfaces.

different directions are the same. An illustration of the volume of negative depth values is given in Figure 16.

In part 1 of this analysis we found that if $\gamma_\epsilon = 0$, the smallest volume of negative depth values is obtained if $\hat{x}_0' = x_0' + \beta_\epsilon' f \frac{(Z_{\max}+Z_{\min})}{2}$, and the intersection of the 0 and $-\infty$ distortion surfaces is at $Z = \frac{Z_{\min}+Z_{\max}}{2}$. In order to derive the position of $\hat{x}_0'$ that minimizes the negative depth values for the general case of $\gamma_\epsilon \neq 0$, we study the change of volume as $\hat{x}_0'$ changes from $x_0' + \beta_\epsilon' f \frac{(Z_{\max}+Z_{\min})}{2}$.

Referring to Figure 17, it can be seen that for any depth value Z, a change in the position of $\hat{x}_0'$ to $\hat{x}_0' + d$, assuming $Z \neq 0$, causes the corresponding area of negative depth values to change by A_c, where

$$A_c = -(y'_1 + y'_2)\, d\, \mathrm{sgn}(\gamma_\epsilon)$$

and y'_1 and y'_2 denote the y' coordinates of the intersection point of the $-\infty$ distortion contour at depth Z with the 0 distortion contours $x' = \hat{x}_0'$ and $x' = \hat{x}_0' + d$.

By intersecting the $-\infty$ distortion contour $x' - x_0' + Z(-\beta_\epsilon' f + \gamma y) = 0$ with the 0 distortion contours $x' - (x_0' + \frac{\beta_\epsilon'}{2}(Z_{\min}+Z_{\max})) = 0$ and $x' - (x_0' + \frac{\beta_\epsilon'}{2}(Z_{\min}+Z_{\max}) + d) = 0$,

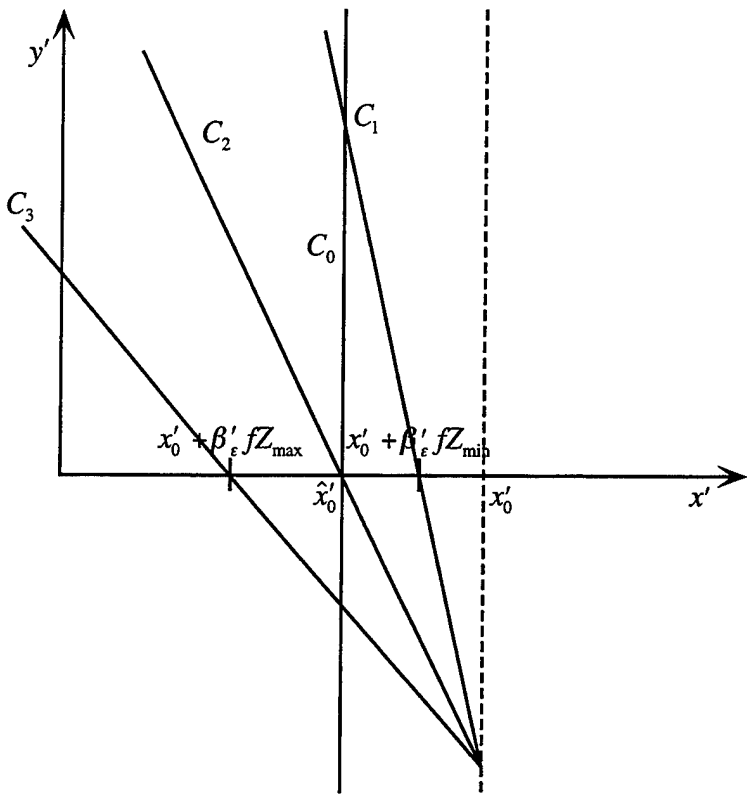


Figure 15: Slices parallel to the $x'y'$ plane through the 0 distortion surface (C_0) and the $-\infty$ distortion surfaces at depth values $Z = Z_{\min}$ (C_1), $Z = -\frac{x'_{0_\epsilon}}{\beta_\epsilon' f}$ (C_2), and $Z = Z_{\max}$ (C_3).

we obtain

$$y'_1 = -\frac{\beta_\epsilon'(Z_{\min}+Z_{\max})}{2Z\gamma_\epsilon} + \frac{\beta_\epsilon'}{\gamma_\epsilon}$$
$$\text{and} \quad y'_2 = -\frac{\beta_\epsilon'}{2}\frac{(Z_{\min}+Z_{\max})}{Z\gamma_\epsilon} - \frac{d}{Z\gamma_\epsilon} + \frac{\beta_\epsilon'}{\gamma_\epsilon}$$

and therefore

$$A_c = -\text{sgn}\,(\gamma_\epsilon)\, d \left(\frac{2\beta_\epsilon'}{\gamma_\epsilon} - \left(\frac{\beta_\epsilon'\,(Z_{\min} + Z_{\max}) + d}{Z\gamma_\epsilon} \right) \right)$$

The change in negative depth volume for any direction V_c is thus given by

$$V_c = \text{sgn}\,(\beta_\epsilon') \int_{x_0'+\beta_\epsilon' Z_{\min}}^{x_0'+\beta_\epsilon' Z_{\max}} A_c$$

which amounts to

$$V_c = \text{sgn}\,(\beta_\epsilon')\,\text{sgn}\,(\gamma_\epsilon)\, d \left(\frac{2\beta_\epsilon'}{\gamma_\epsilon}\,(Z_{\max} - Z_{\min}) - \left(\frac{\beta_\epsilon'\,(Z_{\min} + Z_{\max}) + d}{\gamma_\epsilon} \right) \ln \left(\frac{Z_{\max}}{Z_{\min}} \right) \right)$$

It can be verified that in order for V_c to be negative, $\text{sgn}(\beta_\epsilon') = -\text{sgn}(d)$. This means that $x_0' + d$ lies between $\hat{x}'_0$ and x_0' (see Figure 17).

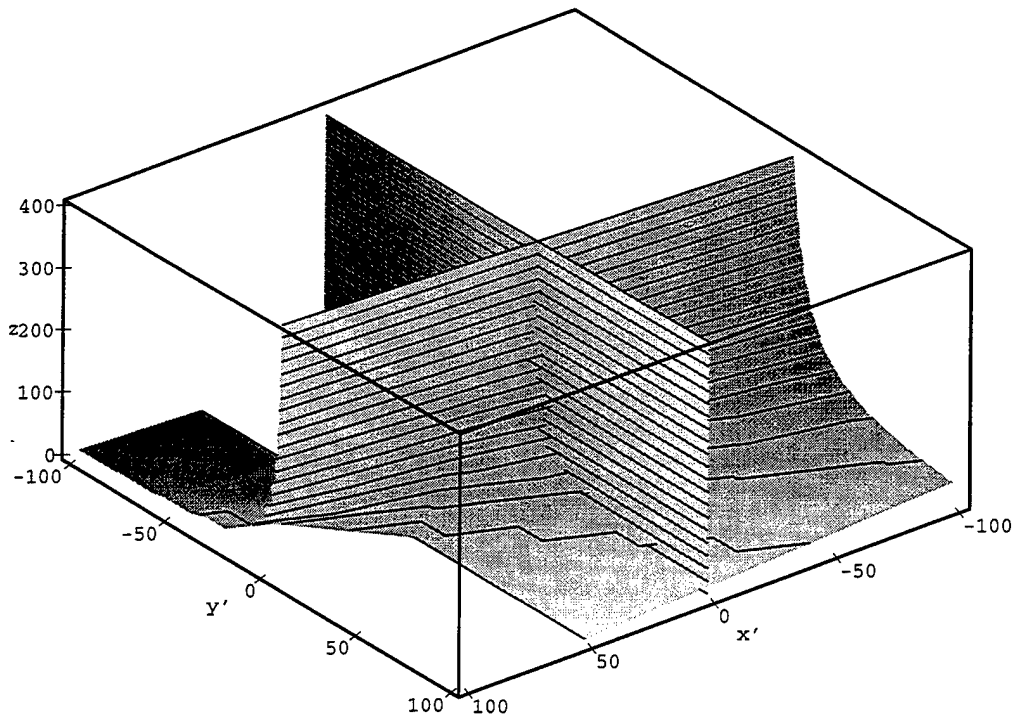


Figure 16: $\gamma_\epsilon \neq 0$: volume of negative depth values between the 0 and $-\infty$ distortion surfaces.

We are interested in the d which minimizes V_c. By solving

$$\frac{\partial V_c}{\partial d} = 0$$

we obtain

$$d = \beta_\epsilon{}' \left[\frac{(Z_{\max} - Z_{\min})}{\ln\left(\frac{Z_{\max}}{Z_{\min}}\right)} - \frac{1}{2}(Z_{\min} + Z_{\max}) \right]$$

The change, Z_c, in the Z coordinate of the intersection between the 0 and $-\infty$ distortion surface is $Z_c = \frac{d}{\beta_\epsilon{}'}$ and thus the intersection for the smallest negative depth volume is at

$$Z_m = \frac{Z_{\max} - Z_{\min}}{\ln\left(\frac{Z_{\max}}{Z_{\min}}\right)} \tag{15}$$

Since Z_m is the same for all flow directions, the total negative depth volume is obtained if the volume in every direction is minimized. Therefore we have the constraint

$$\frac{x_{0_\epsilon}}{y_{0_\epsilon}} = -\frac{\beta_\epsilon}{\alpha_\epsilon} \tag{16}$$

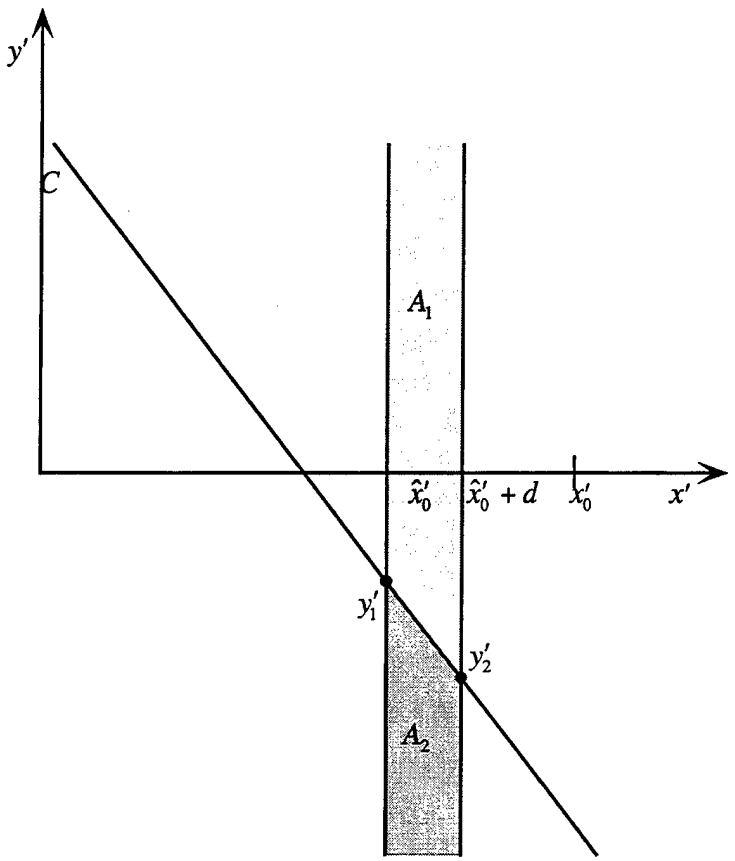


Figure 17: A change of $\hat{x}'_0$ to $\hat{x}'_0 + d$ causes the area of negative depth values A_c to increase by area A_1 and to decrease by area A_2. This change amounts to $A_c = -(y'_1 + y'_2)d\,\mathrm{sgn}(\gamma_\epsilon)$.

For a given rotational error $(\alpha_\epsilon, \beta_\epsilon, \gamma_\epsilon)$, equation (16) defines the direction of the FOE of the translational error on the image plane. For a given translational error $(x_{0_\epsilon}, y_{0_\epsilon})$ equation (16) defines the direction of the AOR of the rotational error on the image. In addition we must have $\gamma_\epsilon = 0$.

Some comment on the finiteness of the image is necessary here. The values A_c and V_c have been derived for an infinitely large image. If γ_ϵ is very small or some of the depth values Z in the interval $[Z_{\min}, Z_{\max}]$ are small, the coordinates of the intersections y'_1 and y'_2 do not lie within the image. The value of A_c can be at most the length of the image times d. Since the slope of the $-\infty$ distortion contour for a given Z is the same for all directions, this will have very little effect on the relationship between the translational and rotational motion errors. It has an effect, however, on the value Z_m.

Assuming the intersections are within the image, we can also derive the relative values of the motion errors: The amount of error depends on the interval of depth values of the scene in view. Since for every direction

$$\hat{x}'_0 = x_0{}' + \beta_\epsilon{}' f Z_m$$

we obtain by substitution from equation (15)

$$x_{0_\epsilon} = -\beta_\epsilon f \left(\frac{Z_{\max} - Z_{\min}}{\ln\left(\frac{Z_{\max}}{Z_{\min}}\right)} \right)$$

$$\text{and} \quad y_{0_\epsilon} = \alpha_\epsilon f \left(\frac{Z_{\max} - Z_{\min}}{\ln\left(\frac{Z_{\max}}{Z_{\min}}\right)} \right)$$

5 Shape Estimation in the Presence of Distortion

The above results are of great importance for the analysis of shape estimation. An error of the form $\frac{x_{0_\epsilon}}{y_{0_\epsilon}} = -\frac{\beta_\epsilon}{\alpha_\epsilon}$ guarantees that for the image near the fixation center, a shape map of the scene is derived which is very well behaved.

Near the image center the image coordinates are very small. Thus using equation (12) the distortion factor there can be approximated by

$$D = \frac{\hat{x}_0 n_x + \hat{y}_0 n_y}{x_0 n_x + y_0 n_y + Zf\left(\beta_\epsilon n_x - \alpha_\epsilon n_y\right)}$$

If $\frac{x_{0_\epsilon}}{y_{0_\epsilon}} = -\frac{\beta_\epsilon}{\alpha_\epsilon}$ for any given Z, the numerator is a multiple of the denominator and thus the distortion factor is the same for every direction (n_x, n_y). This means that scene points of the same depth are distorted by the same factor and the computed depth map has the same level contours as the actual depth map of the scene.

Depending on the sign of the rotational error, there will either be an overestimation for the nearby scene and an underestimation for the far scene or vice versa. All the distortion, however, takes place only in the Z dimension. Thus the resulting depth function involves an affine transformation. The invariants of these shape maps have been studied in the work of Koenderink and van Doorn [22, 23].

6 Conclusions

An algorithm-independent stability analysis of structure from motion has been presented. The analysis did not make any assumptions about the scene, and was based solely on the fact that the depth of the scene—in order for the scene to be visible—has to be positive. As input to the structure from motion process we did not consider optic flow or correspondence, but the value of the flow at every point along some direction, a quantity more easily computable. Our stability analysis amounts to an understanding of the coupling of the translational and rotational error. Given an error in the translation (or the rotation), we asked: what is the value of the rotation (or the translation) that estimates the minimum number of negative depth values? We performed the analysis for both a spherical and a planar retina. For the case of a planar retina we found that the configuration of the rotational and translational errors resulting in minimum negative depth is the one in which the projections of the two error vectors on the image plane are perpendicular to each other. For the case of a spherical retina, we found that given a rotational error, the optimal translation is the correct one, while given an error in

translation, the optimal rotation error is perpendicular to the translational error at an equal distance from the real and estimated translations.

These results, besides their potential use in structure from motion algorithms, also represent a computational analysis comparing different eye constructions in the natural world. The results on the sphere demonstrate that it is very easy for a system with panoramic vision to estimate its self-motion. Indeed, if the system possesses an inertial sensor providing its rotation with some error, we have shown that after derotation, a simple algorithm considering only translation based on normal flow will estimate the translation optimally. This suggests that spherical eye design is optimal for flying systems such as the compound eyes of insects and the panoramic vision of birds.

The analysis on the plane revealed that for an optimal configuration of errors, the estimated depth distorts only in the z direction, with the level contours of the depth function distorting by the same amount, thus making it feasible to extract meaningful shape representations. This suggests that the camera-type eyes of primates are possibly optimal for systems that need good shape computation capabilities.

References

[1] G. Adiv. Inherent ambiguities in recovering 3-D motion and structure from a noisy flow field. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11:477–489, 1989.

[2] Y. Aloimonos and Z. Duric. Estimating the heading direction using normal flow. *International Journal of Computer Vision*, 13:33–56, 1994.

[3] T. Brodsky, C. Fermüller, and Y. Aloimonos. Directions of motion fields are hardly ever ambiguous. Technical Report CAR-TR-780, Center for Automation Research, University of Maryland, 1995.

[4] A. Bruss and B. Horn. Passive navigation. *Computer Vision, Graphics, and Image Processing*, 21:3–20, 1983.

[5] L. Cheong, C. Fermüller, and Y. Aloimonos. Interaction between 3D shape and motion: Theory and applications. Technical Report CS-TR-3480, Center for Automation Research, University of Maryland, June 1996.

[6] K. Daniilidis. *On the Error Sensitivity in the Recovery of Object Descriptions.* PhD thesis, Department of Informatics, University of Karlsruhe, Germany, 1992, in German.

[7] K. Daniilidis and M. Spetsakis. Understanding noise sensitivity in structure from motion. In Y. Aloimonos, editor, *Visual Navigation: From Biological Systems to Unmanned Ground Vehicles*, chapter 4. Lawrence Erlbaum Associates, Hillsdale, NJ, 1996. In press.

[8] O. Faugeras. *Three-Dimensional Computer Vision.* MIT Press, Cambridge, MA, 1992.

[9] O. Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig? In G. Sandini, editor, *Proc. 2nd European Conference on Computer Vision*, pages 563–578, Santa Margherita Ligure, Italy, 1992. Springer-Verlag.

[10] O. Faugeras and B. Mourrain. On the geometry and algebra of the point and line correspondences between n images. In *Proc. 5th International Conference on Computer Vision*, pages 951–956, Cambridge, MA, 1995.

[11] C. Fermüller and Y. Aloimonos. Direct perception of three-dimensional motion from patterns of visual motion. *Science*, 270:1973–1976, 22 December 1995.

[12] C. Fermüller and Y. Aloimonos. Qualitative egomotion. *International Journal of Computer Vision*, 15:7–29, 1995.

[13] C. Fermüller and Y. Aloimonos. On the geometry of visual correspondence. *International Journal of Computer Vision*, 1996. To appear.

[14] C. Fermüller, L. Cheong, and Y. Aloimonos. Explaining human visual space distortion. Technical report, Center for Automation Research, University of Maryland, 1996. To appear.

[15] R. Hartley. Projective reconstruction and invariants from multiple images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16:1036–1041, 1994.

[16] W. Hofmann. Das Problem der "gefährlichen Flächen" in Theorie and Praxis. Deutsche Geodätische Kommission bei der Bayerischen Akademie der Wissenschaften, Reihe C, Heft 3, München, 1953.

[17] B. Horn. *Robot Vision.* McGraw Hill, New York, 1986.

[18] B. Horn. Motion fields are hardly ever ambiguous. *International Journal of Computer Vision*, 1:259–274, 1987.

[19] B. Horn. Relative orientation. *International Journal of Computer Vision*, 4:59–78, 1990.

[20] B. Horn and E. Weldon, Jr. Direct method for recovering motion. *International Journal of Computer Vision*, 2:51–76, 1988.

[21] J. Koenderink and A. van Doorn. Affine structure from motion. *Journal of the Optical Society of America*, 8:377–385, 1991.

[22] J. Koenderink and A. van Doorn. Two-plus-one-dimensional differential geometry. *Pattern Recognition Letters*, 15:439–443, 1994.

[23] J. Koenderink and A. van Doorn. Relief: Pictorial and otherwise. *Image and Vision Computing*, 13:321–334, 1995.

[24] H. Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. *Nature*, 293:133–135, 1981.

[25] H. C. Longuet-Higgins and K. Prazdny. The interpretation of a moving retinal image. *Proceedings of the Royal Society, London B*, 208:385–397, 1980.

[26] S. Maybank. *A Theoretical Study of Optical Flow.* PhD thesis, University of London, UK, November 1987.

[27] S. Maybank. *Theory of Reconstruction from Image Motion.* Springer, Berlin, 1993.

[28] K. Åström. *Invariancy Methods for Points, Curves and Surfaces in Computational Vision.* PhD thesis, Department of Mathematics, Lund Institute of Technology, Lund, Sweden, May 1996.

[29] S. Negahdaripour. *Direct Passive Navigation.* PhD thesis, Department of Mechanical Engineering, MIT, Cambridge, MA, 1986.

[30] R. Nelson and J. Aloimonos. Finding motion parameters from spherical flow fields (or the advantage of having eyes in the back of your head). *Biological Cybernetics*, 58:261–273, 1988.

[31] T. Poggio and W. Reichardt. Considerations on models of movement detection. *Kybernetik*, 13:223–227, 1973.

[32] W. Reichardt. Autocorrelation, a principle for evaluation of sensory information by the central nervous system. In W. Rosenblith, editor, *Principles of Sensory Communication*, pages 303–317. John Wiley and Sons, New York, 1961.

[33] W. Reichardt. Evaluation of optical motion information by movement detectors. *J. Comp. Physiol.*, 161:533–547, 1987.

[34] D. Sinclair, A. Blake, and D. Murray. Robust estimation of egomotion from normal flow. *International Journal of Computer Vision*, 13:57–69, 1994.

[35] C. Slama, C. Theurer, and S. Henriksen. *Manual of Photogrammetry.* American Society of Photogrammetry, Falls Church, VA, 1980.

[36] M. Spetsakis. Models of statistical visual motion estimation. *Computer Vision, Graphics, and Image Processing*, 60:300–312, 1994.

[37] M. Spetsakis and J. Aloimonos. Optimal computing of structure from motion using point correspondence. In *Proc. Second International Conference on Computer Vision*, pages 449–453, Tampa, FL, 1988.

[38] M. Spetsakis and J. Aloimonos. Optimal motion estimation. In *Proc. IEEE Workshop on Visual Motion*, pages 229–237, Irvine, CA, 1989.

[39] M. Spetsakis and J. Aloimonos. Structure from motion using line correspondences. *International Journal of Computer Vision*, 4:171–183, 1990.

[40] M. Spetsakis and J. Aloimonos. A unified theory of structure from motion. In *Proc. DARPA Image Understanding Workshop*, pages 271–283, Pittsburgh, PA, 1990.

[41] R. Tsai and T. Huang. Uniqueness and estimation of three-dimensional motion parameters of rigid objects with curved surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:13–27, 1984.

[42] S. Ullman. *The Interpretation of Visual Motion.* MIT Press, Cambridge, MA, 1979.

[43] J. van Santen and G. Sperling. Temporal covariance model of human motion perception. *Journal of the Optical Society of America A*, 1:451–473, 1984.

[44] A. Waxman and K. Wohn. Contour evolution, neighborhood deformation and global image flow: Planar surfaces in motion. *International Journal of Robotics Research*, 4(3):95–108, 1985.

Appendix A Re-parameterization of Flow Directions

Let us choose a uniformly distributed flow field direction $\mathbf{n}_1(\psi)$ as follows. The coordinates of $\mathbf{r} = [x, y, z]^T$ at every point on the unit sphere are obtained through a rotation of point $[0, 0, 1]^T$ by an angle φ_x around the x axis followed by a rotation of angle φ_y around the y axis. Thus the rotation matrix R is given by

$$R = \begin{bmatrix} \cos\varphi_y & 0 & \sin\varphi_y \\ -\sin\varphi_x \sin\varphi_y & \cos\varphi_x & \sin\varphi_x \cos\varphi_y \\ -\cos\varphi_x \sin\varphi_y & -\sin\varphi_x & \cos\varphi_x \cos\varphi_y \end{bmatrix}$$

and every point $\mathbf{r} = [\sin\varphi_y, \sin\varphi_x \cos\varphi_y, \cos\varphi_x \cos\varphi_y]^T$.

Vectors $\mathbf{n}_1(\psi)$ are obtained through rotation of unit vector $[\sin\psi, \cos\psi, 0]^T$ at point $[0, 0, 1]^T$. Thus

$$\mathbf{n}_1(\psi) = [\cos\varphi_y \sin\psi, -\sin\varphi_x \sin\varphi_y \sin\psi + \cos\varphi_x \cos\psi, -\cos\varphi_x \sin\varphi_y \sin\psi - \sin\varphi_x \cos\psi]^T$$

On the other hand, the direction $\mathbf{n}_2(\chi)$ used in the analysis in Section 3.1 is chosen to be $\mathbf{n}_2(\chi) = \mathbf{r} \times \mathbf{s}(\chi)$ with $\mathbf{s} = [0, \sin\chi, \cos\chi]^T$.

Thus $\mathbf{n}_2(\chi) = [\cos\varphi_x \cos\varphi_y \sin\chi - \sin\varphi_x \cos\varphi_y \cos\chi, \sin\varphi_y \cos\chi, -\sin\varphi_y \sin\chi]^T$. In order for $\mathbf{n}_1(\psi)$ to be parallel to $\mathbf{n}_2(\chi)$ the following must hold:

$$(\mathbf{n}_1(\psi) \times \mathbf{n}_2(\chi)) \cdot \mathbf{r} = 0$$

Thus $\psi = g(\chi) = \arctan(\frac{\tan(\chi - \varphi_x)}{\sin\varphi_y})$ and the normalization factor $|\frac{\partial\psi}{\partial\chi}|$ is

$$\left|\frac{\partial\psi}{\partial\chi}\right| = \left|\frac{\sin(\varphi_y)}{\cos(\varphi_y)^2 \cos(\chi - \varphi_x)^2 - 1}\right|$$

For an illustration see Figure 6.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	September 1996	Technical Report

4. TITLE AND SUBTITLE

Algorithm-Independent Stability Analysis of Structure from Motion

5. FUNDING NUMBERS

N00014-96-1-0587

6. AUTHOR(S)

C. Fermüller and Y. Aloimonos

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Computer Vision Laboratory
Center for Automation Research
University of Maryland
College Park, MD 20742-3275

8. PERFORMING ORGANIZATION REPORT NUMBER

CAR-TR-840
CS-TR-3691

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Office of Naval Research
800 North Quincy Street
Arlington, VA 22203-1714

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Office of Naval Research position, policy, or decision, unless so designated by other documentation.

12a. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release.
Distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

The stability analysis for the structure from motion problem presented in this paper investigates the optimal relationship between the errors in the estimated translational and rotational parameters of a rigid motion that results in the estimation of a minimum number of negative depth values. No particular estimators are used and no specific assumptions about the scene are made. The input used is the value of the flow along some direction, which is more general than optic flow or correspondence. For a planar retina it is shown that the optimal configuration is achieved when the projections of the translational and rotational errors on the image plane are perpendicular. For a spherical retina, given a rotational error, the optimal translation is the correct one, while given a translational error the optimal rotational error is normal to the translational one at an equal distance from the real and estimated translations. The proofs, besides illuminating the confounding of translation and rotation in structure from motion, have an important application to ecological optics. The same analysis provides a computational explanation of why it is much easier to estimate self-motion in the case of a spherical retina and why it is much easier to estimate shape in the case of a planar retina, thus suggesting that nature's design of compound eyes (or panoramic vision) for flying systems and camera-type eyes for primates (and other systems that perform manipulation) is optimal.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Structure from motion, stability analysis, spherical retina, ecological optics	36
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

GENERAL INSTRUCTIONS FOR COMPLETING SF 298

The Report Documentation Page (RDP) is used in announcing and cataloging reports. It is important that this information be consistent with the rest of the report, particularly the cover and title page. Instructions for filling in each block of the form follow. It is important to ***stay within the lines*** to meet ***optical scanning requirements.***

Block 1. Agency Use Only *(Leave blank)*.

Block 2. Report Date. Full publication date including day, month, and year, if available (e.g. 1 Jan 88). Must cite at least the year.

Block 3. Type of Report and Dates Covered. State whether report is interim, final, etc. If applicable, enter inclusive report dates (e.g. 10 Jun 87 - 30 Jun 88).

Block 4. Title and Subtitle. A title is taken from the part of the report that provides the most meaningful and complete information. When a report is prepared in more than one volume, repeat the primary title, add volume number, and include subtitle for the specific volume. On classified documents enter the title classification in parentheses.

Block 5. Funding Numbers. To include contract and grant numbers; may include program element number(s), project number(s), task number(s), and work unit number(s). Use the following labels:

C	- Contract	PR	- Project
G	- Grant	TA	- Task
PE	- Program Element	WU	- Work Unit Accession No.

Block 6. Author(s). Name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. If editor or compiler, this should follow the name(s).

Block 7. Performing Organization Name(s) and Address(es). Self-explanatory.

Block 8. Performing Organization Report Number. Enter the unique alphanumeric report number(s) assigned by the organization performing the report.

Block 9. Sponsoring/Monitoring Agency Name(s) and Address(es). Self-explanatory.

Block 10. Sponsoring/Monitoring Agency Report Number. *(If known)*

Block 11. Supplementary Notes. Enter information not included elsewhere such as: Prepared in cooperation with...; Trans. of...; To be published in.... When a report is revised, include a statement whether the new report supersedes or supplements the older report.

Block 12a. Distribution/Availability Statement. Denotes public availability or limitations. Cite any availability to the public. Enter additional limitations or special markings in all capitals (e.g. NOFORN, REL, ITAR).

DOD	- See DoDD 5230.24, "Distribution Statements on Technical Documents."
DOE	- See authorities.
NASA	- See Handbook NHB 2200.2.
NTIS	- Leave blank.

Block 12b. Distribution Code.

DOD	- Leave blank.
DOE	- Enter DOE distribution categories from the Standard Distribution for Unclassified Scientific and Technical Reports.
NASA	- Leave blank.
NTIS	- Leave blank.

Block 13. Abstract. Include a brief *(Maximum 200 words)* factual summary of the most significant information contained in the report.

Block 14. Subject Terms. Keywords or phrases identifying major subjects in the report.

Block 15. Number of Pages. Enter the total number of pages.

Block 16. Price Code. Enter appropriate price code *(NTIS only)*.

Blocks 17. - 19. Security Classifications. Self-explanatory. Enter U.S. Security Classification in accordance with U.S. Security Regulations (i.e., UNCLASSIFIED). If form contains classified information, stamp classification on the top and bottom of the page.

Block 20. Limitation of Abstract. This block must be completed to assign a limitation to the abstract. Enter either UL (unlimited) or SAR (same as report). An entry in this block is necessary if the abstract is to be limited. If blank, the abstract is assumed to be unlimited.